Journal of
Experimental Psychology

VOL. VII., No. 6 December, 1924

THE EFFECT OF ATTENTION OR MENTAL ACTIVITY ON THE PATELLAR TENDON REFLEX

BY W. W. TUTTLE

From the Laboratories of Physiology, Ohio State University

In studying the knee-jerk Lombard (1) found in one subject that whatever increased or decreased the activity of the central nervous system as a whole increased or diminished the activity of this reflex. Howell (2) states that this general fact is supported, especially as regards mental activity—such, for instance, as the condition of the nervous center controlling the bladder.

In view of the paucity of quantitative data on this subject it was decided to measure the knee-jerk under active and passive conditions of highly socialized forms of peripheral stimulation. We did not feel free to hazard a definition of what psychologists have generally agreed to include under the term 'attention' or 'mental activity' but preferred to assign to the subject a problem which we felt reasonably sure would release the type of response which most psychologists would agree required 'attention' or was performed under 'attentive attitudes.' The terms *passive* and *active* which we have chosen in place of *non-attentive* and *attentive* are not meant to carry theoretical implications as to their psychological character beyond those which the reader himself wishes to interpret from the nature of the problems that were assigned to the subjects.

Data were collected from seven normal subjects in good health, four of whom were well trained in knee-jerk experi-

ments on a special form of apparatus described in detail elsewhere (3). This apparatus automatically administers uniform blows upon the patellar tendon at a constant rate. In order to accomplish this, the speed of a 1750 R.P.M. constant speed motor is reduced to 7 R.P.M. by means of a series of reducing pulleys. An eccentric placed upon the shaft of the last reducing pulley is connected to a sliding carriage. A rod attached to the carriage engages a freely swinging hammer, pulls it backward and automatically releases it to fall against the patellar tendon. The subject is seated in a barber's chair from which the foot rest has been removed. The chair unaltered furnishes up, down and lateral adjustments. Forward and backward movements are made possible by mounting the chair upon a roller base. The necessary tension for the patellar tendon is secured by elevating the thigh a few degrees above the horizontal and holding the leg slightly back of the vertical by a spring which forms part of the recording device. The distal end of the thigh rests in a comfortable support at the end of which is a U-shaped guide so constructed that the width of the U and the angle of its sides can be adjusted.

The record is made by attaching a piece of piano wire to the heel of the subject's shoe and thence to the wheel of an axle. A second piece of wire runs from the axle of the wheel through guides to a stylus suspended from a steel spring which writes on a smoked drum. The wheel and axle serve to reduce the excursion of the stylus so that the record of a hyperirritable subject can be written upon an ordinary kymograph. The records read from right to left due to the fact that the base line is at the top of the record. This method was adopted in order to eliminate extra guide pulleys.

The forward component of the excursion of the foot is used as the index of the patellar tendon reflex movement, and it is this distance which is recorded by the stylus. This is referred to as the 'height.' The ratio of the distance through which the foot moves to that through which the stylus moves is as 4 : 1. In the Tables herein presented this correction has been made. Throughout the experiments the

frequency of stimulation was seven stimuli per minute, all of equal intensity. In securing the average height of the kick for a given interval the value of zero was assigned for the height in all cases in which there was no visible response to the stimulus stroke.

The usual method of procedure was to determine the average 'height' of the knee-jerk of a subject during a 'passive' and an 'active' period. During the passive period the subject sat as quietly as possible in the chair, with his eyes closed. When the kicks were sufficient in number to establish a relatively uniform record he was given a problem which presumably required his attention. The effect of this problem-solving on the knee-jerk was indicated by a change in the height or the uniformity of the record. When the problem was finished the subject again entered upon the passive conditions and the stimuli were continued until the kicks again registered relatively uniform tonicity. In a number of records a 'conversation period' occurred which has been recorded separately in order not to confuse the results with the passive and active periods.

Subject 1 is a man 24 years of age, a junior in the College of Medicine and has made normal progress. Fig. 1 is the kymograph record of his knee-jerk during three passive and two active periods.

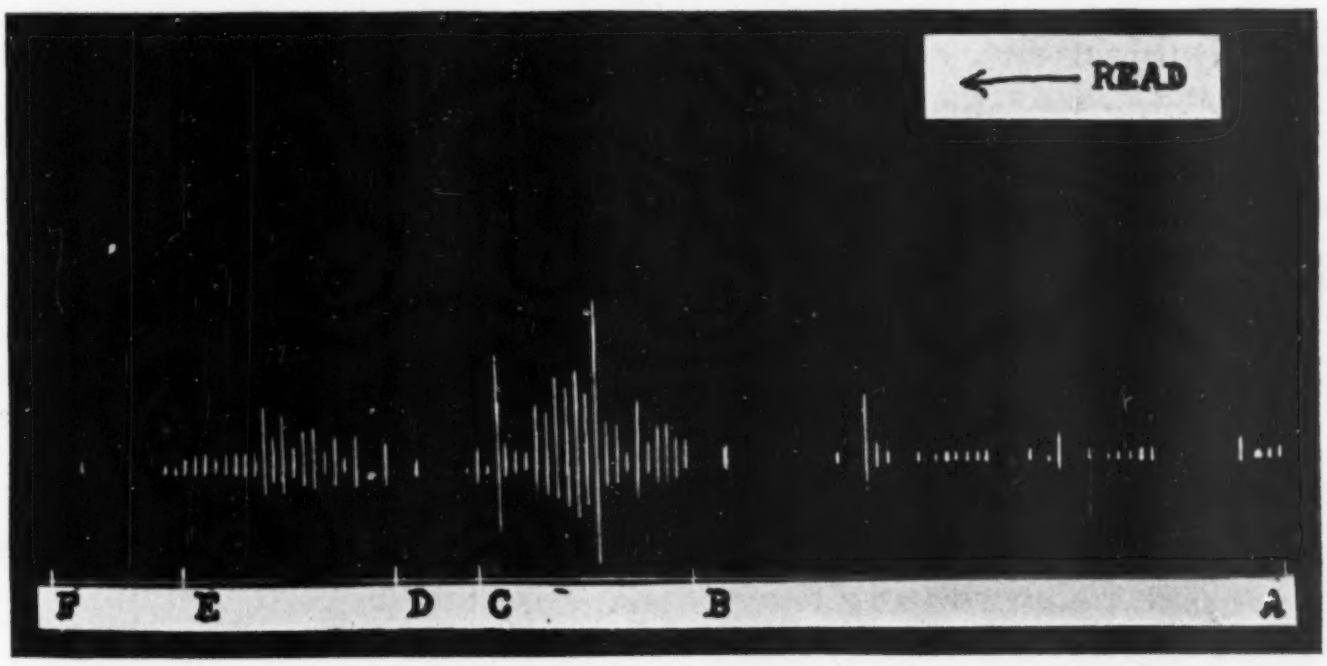

FIG. 1. *A–B*, Passive Period; *B–C*, Problem in Addition; *C–D*, Passive Period; *D–E*, Problem in Square Root; *E–F*, Passive Period

After delivering passive kicks during *A–B* (Fig. 1) the subject was given the following problem: "Find the sum of 8769, 5873, 6798, 9256, 6529, 87879, 478, 697, 1789, 8899, 6784, 3329." This he solved correctly with pencil and paper during the interval *B–C*. This was followed by the passive period *C–D*. Next he attempted, but failed, during the interval *D–E* to extract the square root of 68594386. This was followed by another passive period *E–F*. The analysis of the record as given in Table I. shows that the average height of the kick during the active periods is 10.13 times the average height of the passive periods.

TABLE I

Interval	Time	Total Height of Kicks in Mm.	Number of Stimuli	Average Height per Kick	
				Active	Passive
A–B	7′ 9″	121	50		2.42
B–C	3′ 0″	458	21	21.80	
C–D	1′ 26″	15	10		1.50
D–E	2′ 51″	208	20	10.40	
E–F	1′ 51″	11	13		0.84
			Total	32.20	4.76
			Average	16.10	1.59

16.10/1.59 = 10.13, active/passive, ratio

Subject 2 is a man of 25 years, a junior in the College of Medicine who has made normal progress. Fig. 2 is a record of his knee-jerk during two passive and one active periods.

This record represents the performance of a hypersensitive subject. After the passive interval *A–B* (Fig. 2) the subject attempted to extract the square root of 102535376 during the active period *B–C*, with the use of pencil and paper. The final passive period followed as indicated between *C–D*. The analysis of the record as given in Table II. shows that the average height during the active period is 2.40 times the average height during the passive periods.

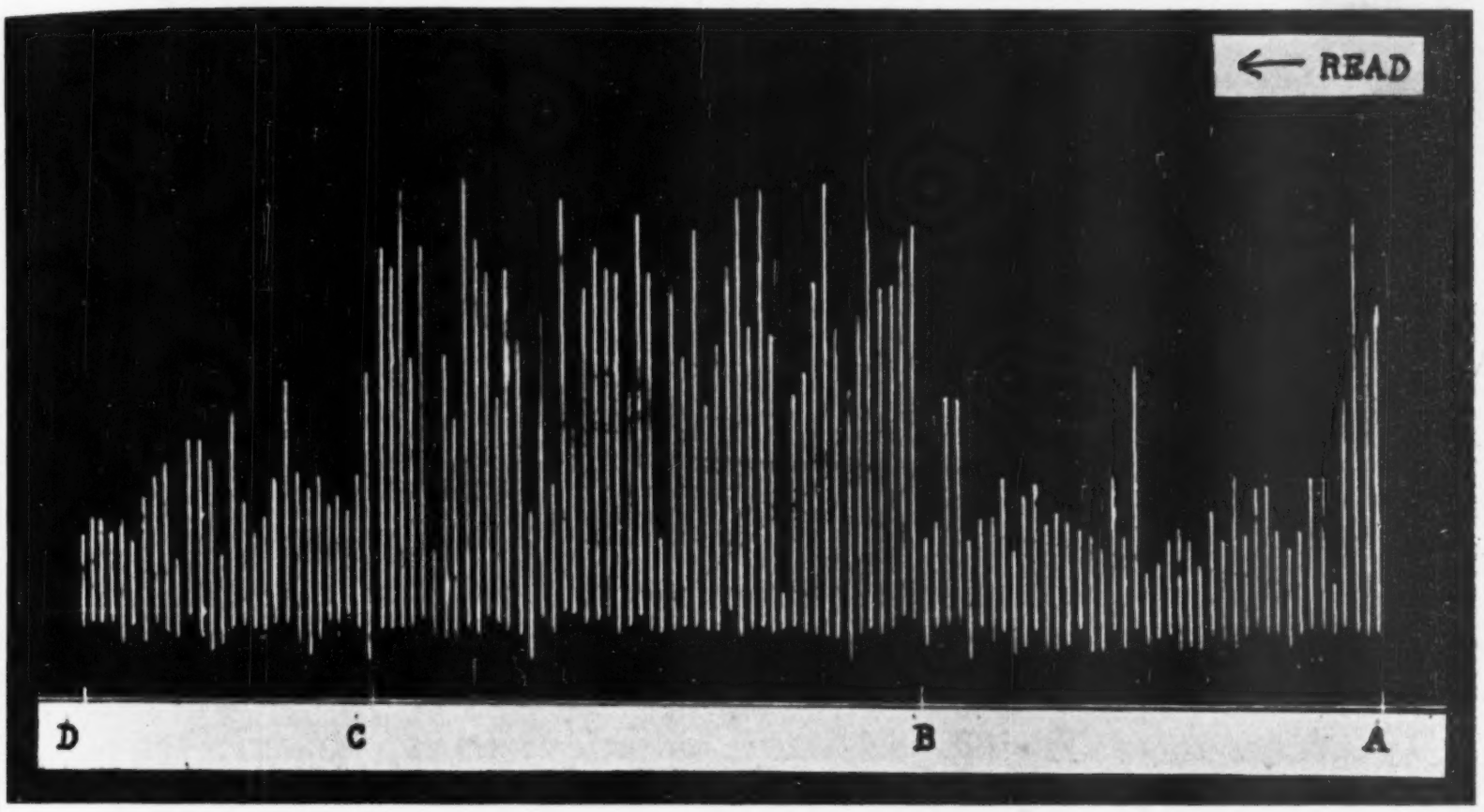

FIG. 2. *A–B*, Passive Period; *B–C*, Problem in Square Root; *C–D*, Passive Period

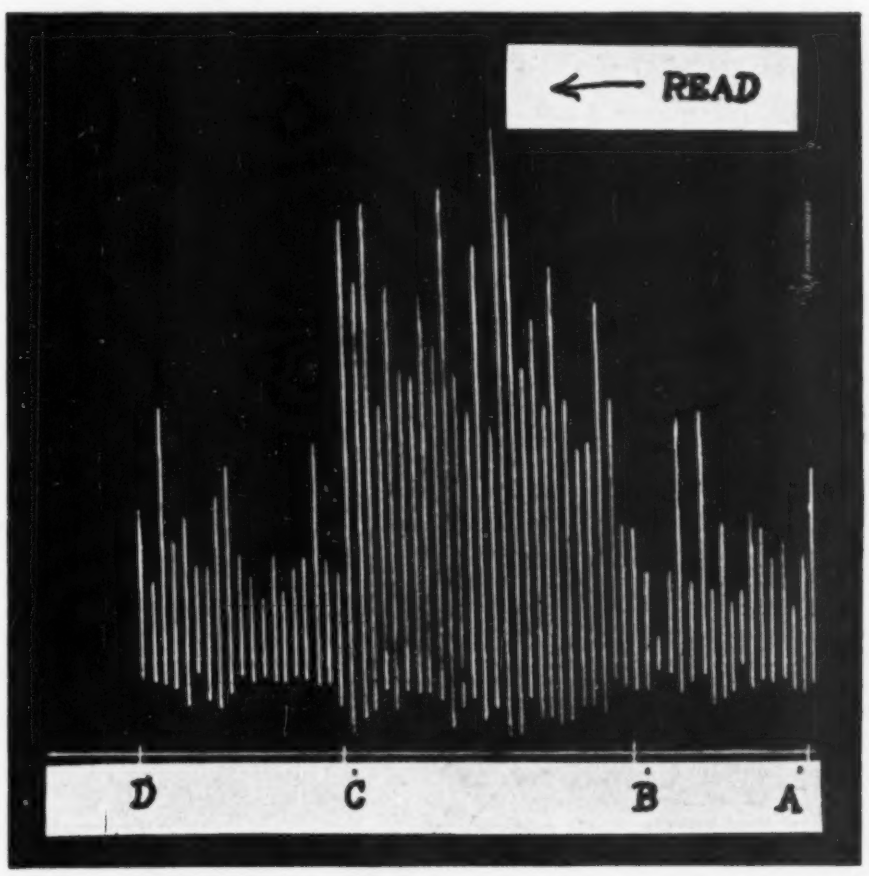

FIG. 2*a*. *A–B*, Passive Period; *B–C*, Problem in Addition; *C–D*, Passive Period

TABLE II

Interval	Time	Total Height of Kicks in Mm.	Number of Stimuli	Average Height per Kick	
				Active	Passive
A–B.....	6′ 0″	2175	42		51.79
B–C.....	7′ 9″	6227	50	124.54	
C–D.....	4′ 0″	1449	28		51.75
			Total.....		103.54
			Average.....	124.54	51.77

124.54/51.77 = 2.40, active/passive, ratio

Another record from the same subject is shown in Fig. 2*a*. The initial passive period occurs between *A–B* (Fig. 2*a*). This is followed by an attempt, indicated between *B–C*, to solve the following problem in addition: 13, 17, 19, 21, 13, 17, 19, 21, 55, 88, 67, 33, 19, 13. The numbers were read to the subject and he performed the cumulative addition orally. This task was followed by the passive period *C–D*. The details of the record as given in Table II., *a* show that the average height during the active period is 2.73 times the average height during the passive periods.

TABLE II., *a*

Interval	Time	Total Height of Kicks in Mm.	Number of Stimuli	Average Height per Kick	
				Active	Passive
A–B.....	2′ 26″	748	17		44.00
B–C.....	3′ 51″	3383	27	125.30	
C–D.....	2′ 43″	909	19		47.84
			Total.....		91.84
			Average.....	125.30	45.92

125.30/45.92 = 2.73, active/passive, ratio

Subject 3 is a colored man of 50 years. Although he spent only two months in school he is able to read, write and solve very simple problems in arithmetic. Fig. 3 is a record of his knee-jerk during three passive and three active periods.

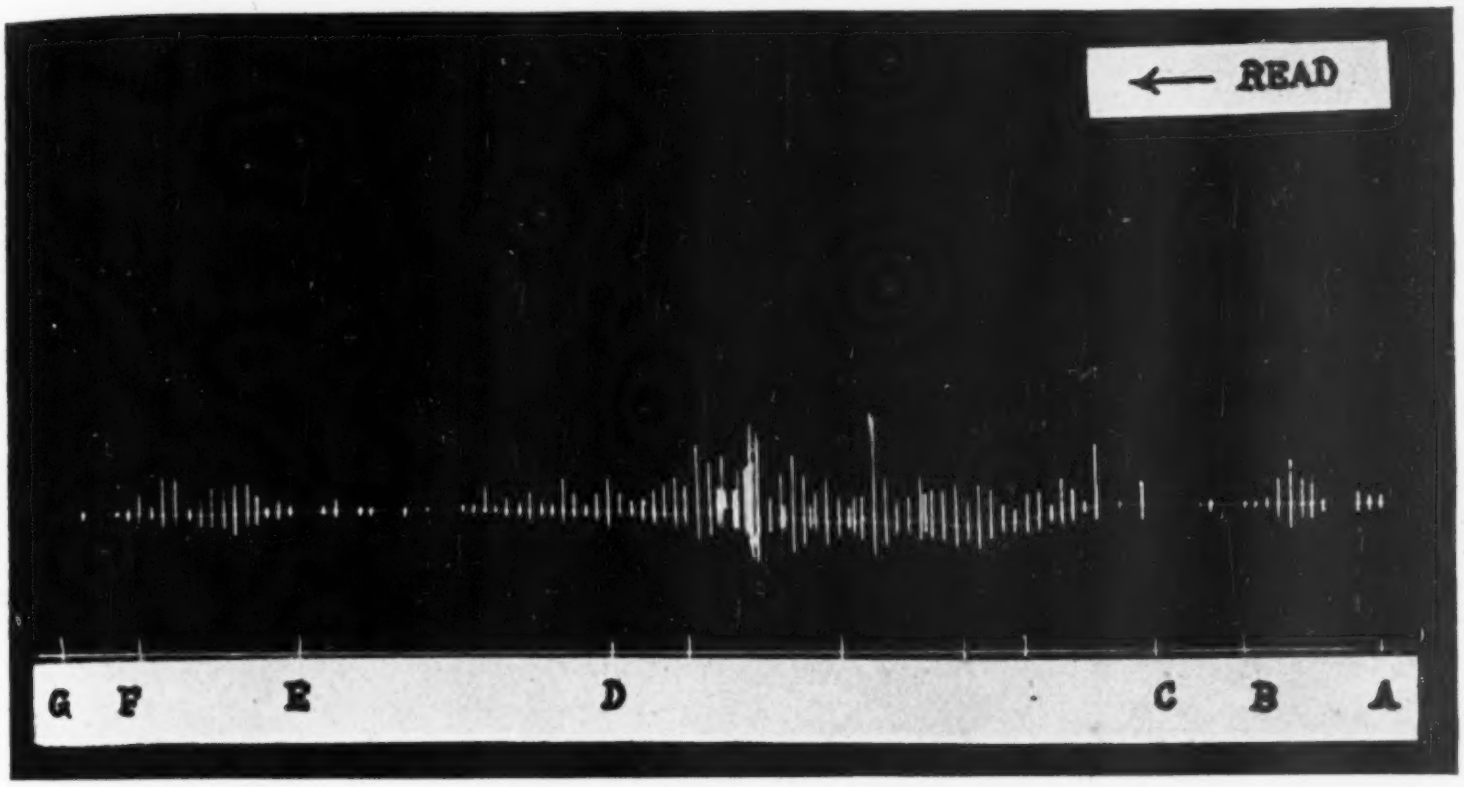

FIG. 3. *A–B*, Received Instructions; *B–C*, Passive Period; *C–D*, Problems in Addition; *D–E*, Passive Period; *E–F*, Gave Personal History; *F–G*, Passive Period

After being told the nature of the experiment during the interval *A–B* (Fig. 3) (recorded as an active period) this subject delivered passive kicks during the period *B–C*. Between *C* and *D* he attempted to solve the following five problems in addition, with the use of pencil and paper:

(*a*) 2, 6, 3, 11, 10
(*b*) 4, 5, 1, 8, 3
(*c*) 8, 7, 14, 21, 6
(*d*) 148, 368, 921, 88, 67
(*e*) 88, 97, 66, 43, 81

Problems *a*, *b*, *c* and *e* were solved correctly. After the problems there followed the passive condition shown between *D* and *E*. During the period *E–F* (recorded as active) he answered questions and reported on his personal history. This was followed by the passive interval *F–G*. For this subject, learning the instructions and reporting his history seemed to require so much effort that they were included as active periods. The details of this record are shown in Table III. The average height of the kick during the active periods was 10.25 times that of the passive periods.

Table III

Interval	Time	Total Height of Kicks in Mm.	Number of Stimuli	Average Height per Kick	
				Active	Passive
A–B	1′ 43″	84	12	7.00	
B–C	1′ 17″	4	9		0.44
C–D	7′ 9″	730	50	14.60	
D–E	4′ 9″	73	29		2.52
E–F	2′ 0″	124	14	8.86	
F–G	1′ 0″	0	7		0.00
			Total	30.46	2.96
			Average	10.15	0.99

10.15/0.99 = 10.25, active/passive, ratio

Subject 4 is a man of 29 years. He is a junior in the College of Medicine and has made normal progress. Fig. 4 is a record of his knee-jerk during four passive and three active periods.

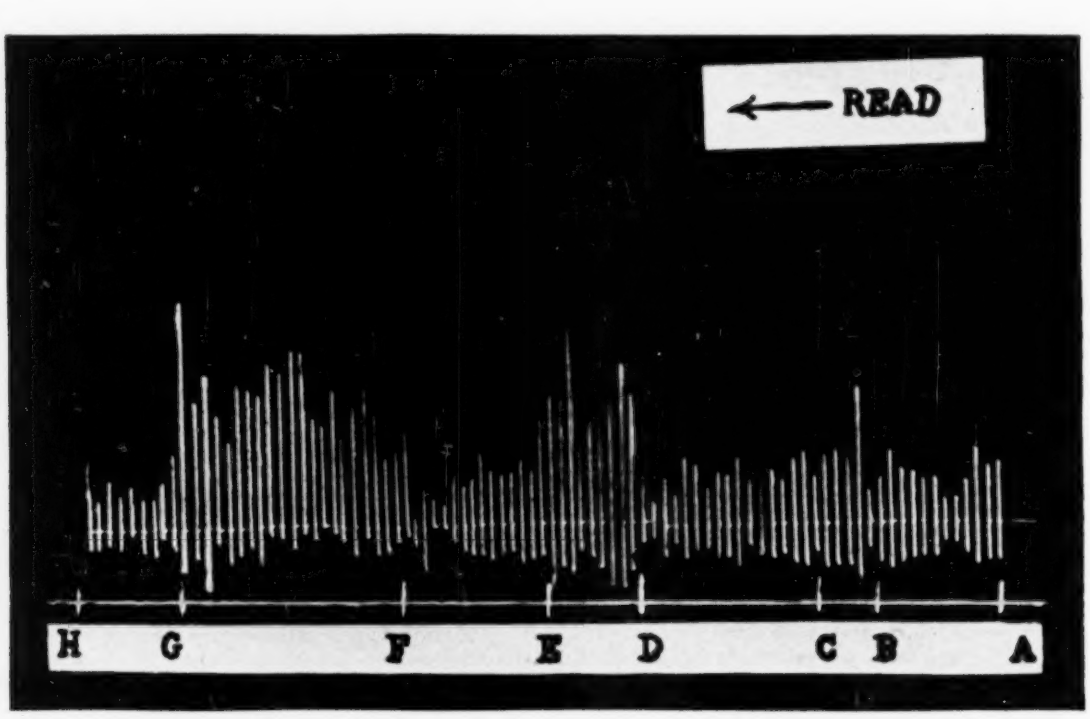

Fig. 4. *A–B*, Passive Period; *B–C*, Instructions were Given; *C–D*, Passive Period; *D–E*, Problem in Addition; *E–F*, Passive Period; *F–G*, Problem in Square Root; *G–H*, Passive Period

Beginning with a passive period during *A–B* (Fig. 4) the subject was instructed as to the nature of his problem during *B–C*. This was followed by a passive period *C–D*. He then solved correctly the following five problems in addition during *D–E*:

(*a*) 9, 11, 21, 76, 18
(*b*) 37, 88, 16, 5, 9
(*c*) 33, 11, 86, 77, 14
(*d*) 21, 36, 83, 41, 60
(*e*) 66, 72, 89, 31, 20.

These problems were followed by a passive period *E–F*. He next attempted to solve, without success, the square root of 368794 during the interval *F–G*. This was followed by a passive period *G–H*. Table IV. gives in detail the data collected from this subject. From a consideration of these results we find that the average height of the kicks during the active periods is 2.22 times the average height during the passive periods.

TABLE IV

Interval	Time	Total Height of Kicks in Mm.	Number of Stimuli	Average Height per Kick	
				Active	Passive
A–B	1′ 43″	255	12		21.25
B–C	0′ 43″	168	5	33.60	
C–D	2′ 26″	354	17		20.82
D–E	1′ 17″	464	9	51.56	
E–F	2′ 0″	299	14		21.36
F–G	3′ 9″	1111	22	50.50	
G–H	1′ 17″	161	9		17.89
			Total	135.66	81.32
			Average	45.22	20.33

45.22/20.33 = 2.22, active/passive, ratio

Subject 5 is a man 22 years of age. He is an instructor in the Department of Physiology. Fig. 5 is a record of his knee-jerk during two passive and one active periods.

After delivering passive kicks during the interval *A–B* (Fig. 5) this subject was given an intelligence test devised by the writer in collaboration with S. L. Pressey (4). The test was completed during the period *B–C*. This was followed by the passive period *C–D*. Table V. gives the details of the data which were collected from this subject. We find that the average height of the kicks during the active period is 3.31 times the average height of the passive periods.

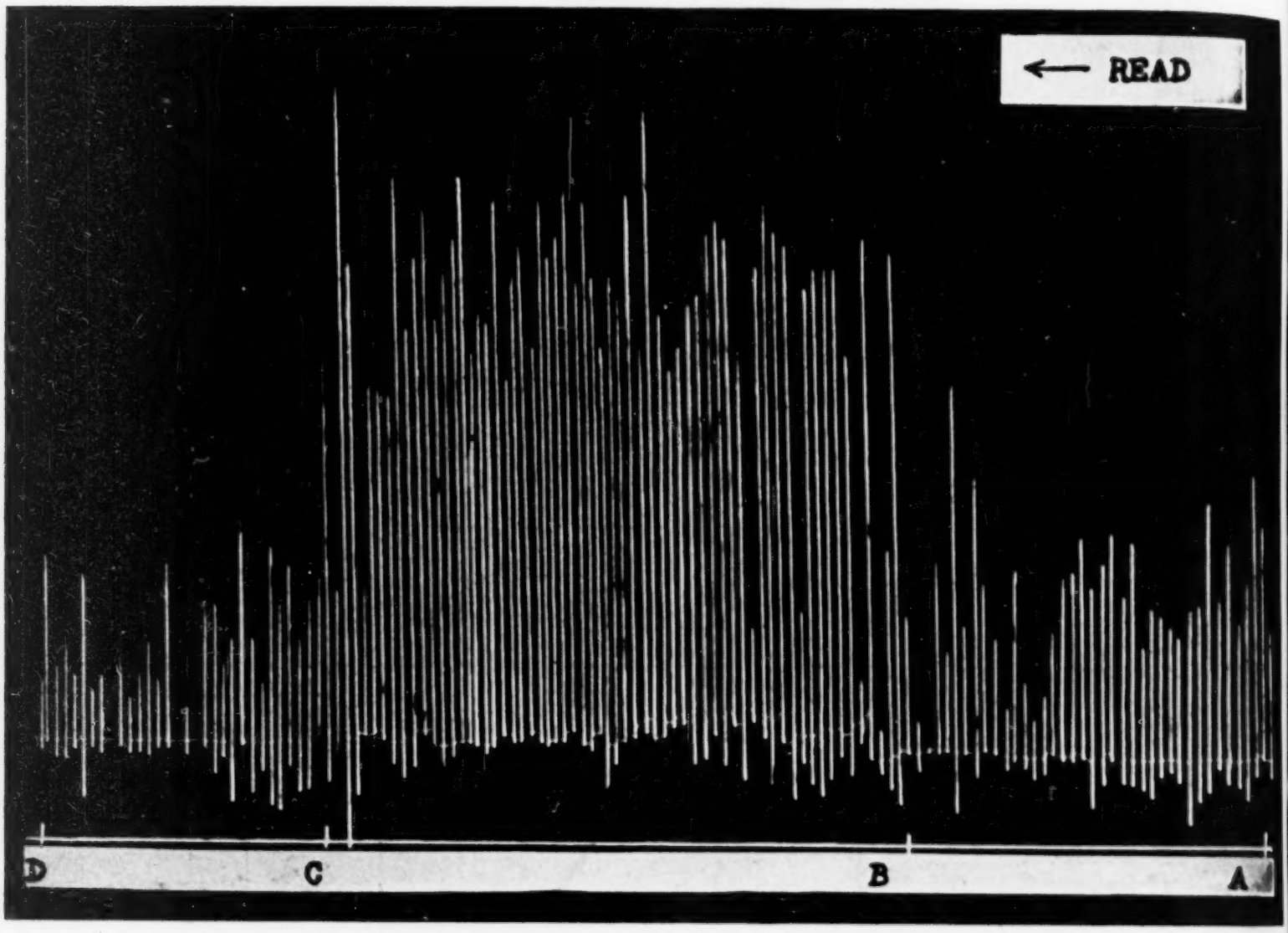

FIG. 5. *A–B*, Passive Period; *B–C*, Intelligence Test; *C–D*, Passive Period

TABLE V

Interval	Time	Total Height of Kicks in Mm.	Number of Stimuli	Average Height per Kick	
				Active	Passive
A–B.....	5′ 34″	2778	39		71.23
B–C.....	8′ 52″	12176	62	196.38	
C–D.....	4′ 26″	1467	31		47.33
			Total.....		118.56
			Average.....	196.38	59.28

196.38/59.28 = 3.31, active/passive, ratio

Subject 6 is a man 31 years of age. He is a student in the Graduate School. Fig. 6 is a record of his knee-jerk during two passive and one active periods.

After delivering passive kicks during the interval *A–B*

(Fig. 6) this subject attempted the following problem in progressive mental addition during *B–C*: 7, 8, 17, 9, 27, 74, 19, 17, 73, 15, 17, 13. When the problem was completed he sat as passive as possible during the interval *C–D*. Table VI. is a detailed report of the record, and shows that the average height of the kicks during the active period is 3.91 times the average height of the passive periods.

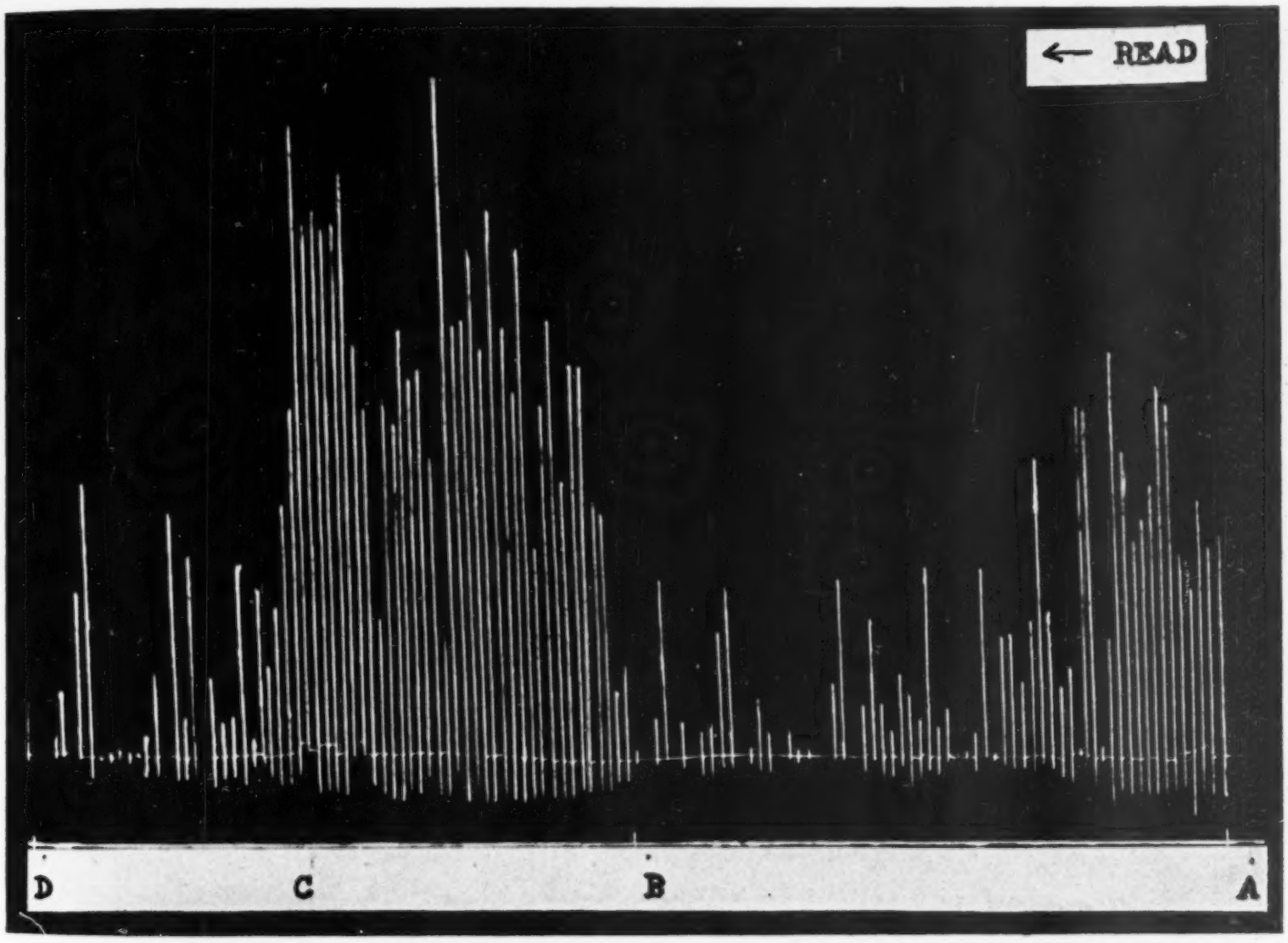

FIG. 6. *A–B*, Passive Period; *B–C*, Problem in Oral Addition; *C–D*, Passive Period. This graph shows a feature (*A–B*) which is quite characteristic of knee-jerk records in the case of both trained and untrained subjects. The initial high kicks followed by a rather gradual fall happens during the transition of the subject from an active to a passive state.

Another record from Subject 6 is given in Fig. 6*a*.

After delivering passive kicks during *A–B* (Fig. 6*a*) the subject attempted to extract the square root of 102535376 during the interval *B–C*. This was followed by a passive period *C–D*. Table VI., *a* shows the details of this record.

It shows that the average height of the kicks during the active period is 10.25 times the average height of the passive periods.

TABLE VI

Interval	Time	Total Height of Kicks in Mm.	Number of Stimuli	Average Height per Kick	
				Active	Passive
A–B.....	9′ 0″	3215	63	177.80	51.03
B–C.....	5′ 0″	6223	35		
C–D.....	4′ 17″	1193	30		39.76
			Total.....	177.80	90.79
			Average.....		45.39

177.80/45.39 = 3.91, active/passive, ratio

The two records of this subject show a variation in the active-passive ratios from 3.91 to 10.25. This may be due to

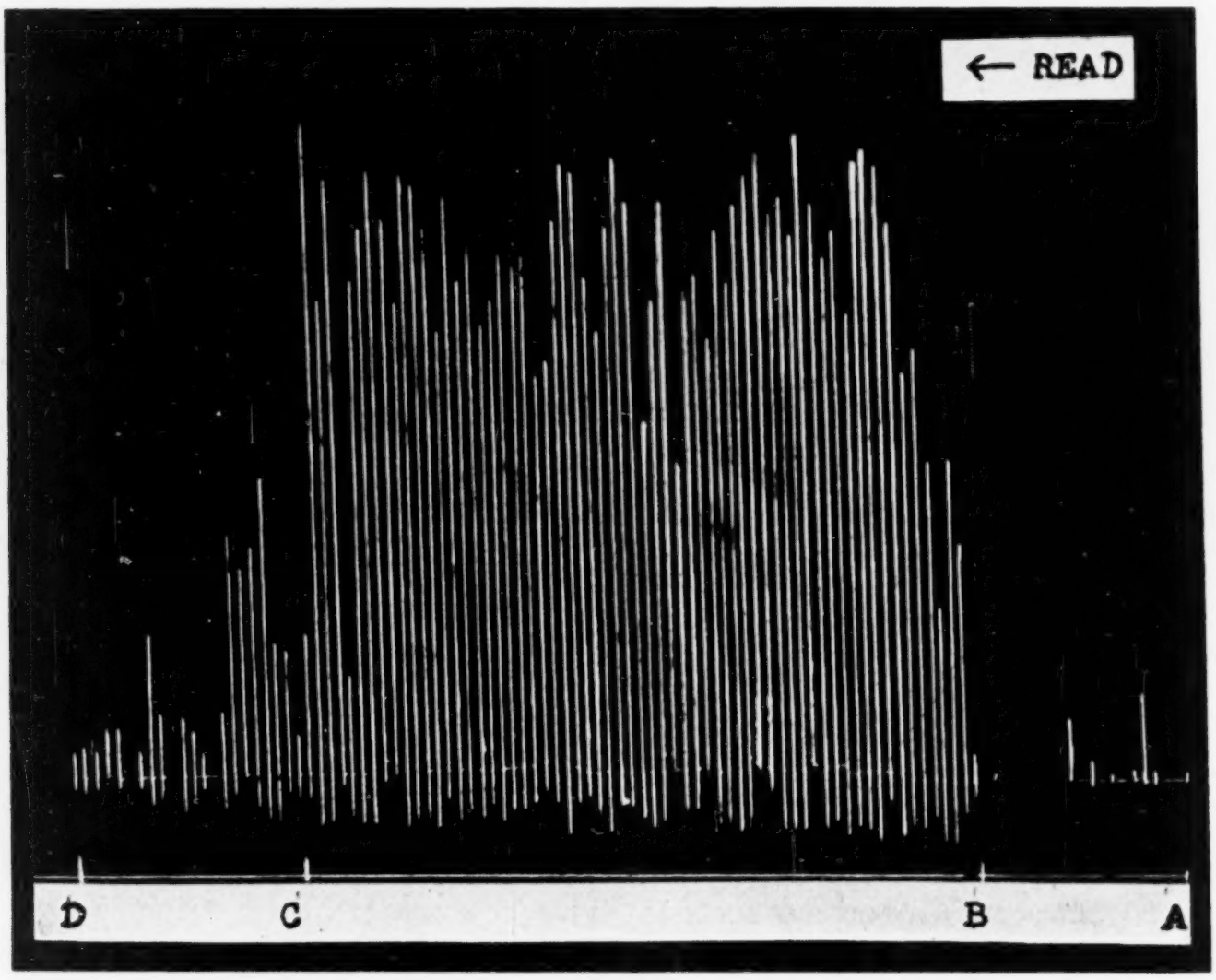

FIG. 6*a*. *A–B*, Passive Period; *B–C*, Problem in Square Root; *C–D*, Passive Period

TABLE VI., *a*

Interval	Time	Total Height of Kicks in Mm.	Number of Stimuli	Average Height per Kick	
				Active	Passive
A–B.....	2′ 51″	91	20		4·55
B–C.....	9′ 9″	12443	64	194.42	
C–D.....	3′ 0″	701	21		33.38
			Total.....		37.93
			Average.....	194.42	18.96

194.42/18.96 = 10.25, active/passive, ratio

the difference in the difficulties of the two problems; mental arithmetic for the lower ratio, a pencil and paper square root problem for the higher ratio.

Subject 7 is a man of 26 years. He has completed three years of college work. He has unusual mechanical ability and has been well trained in knee-jerk experiments. Fig. 7 is a record of his knee-jerk during a conversation and an active or problem period.

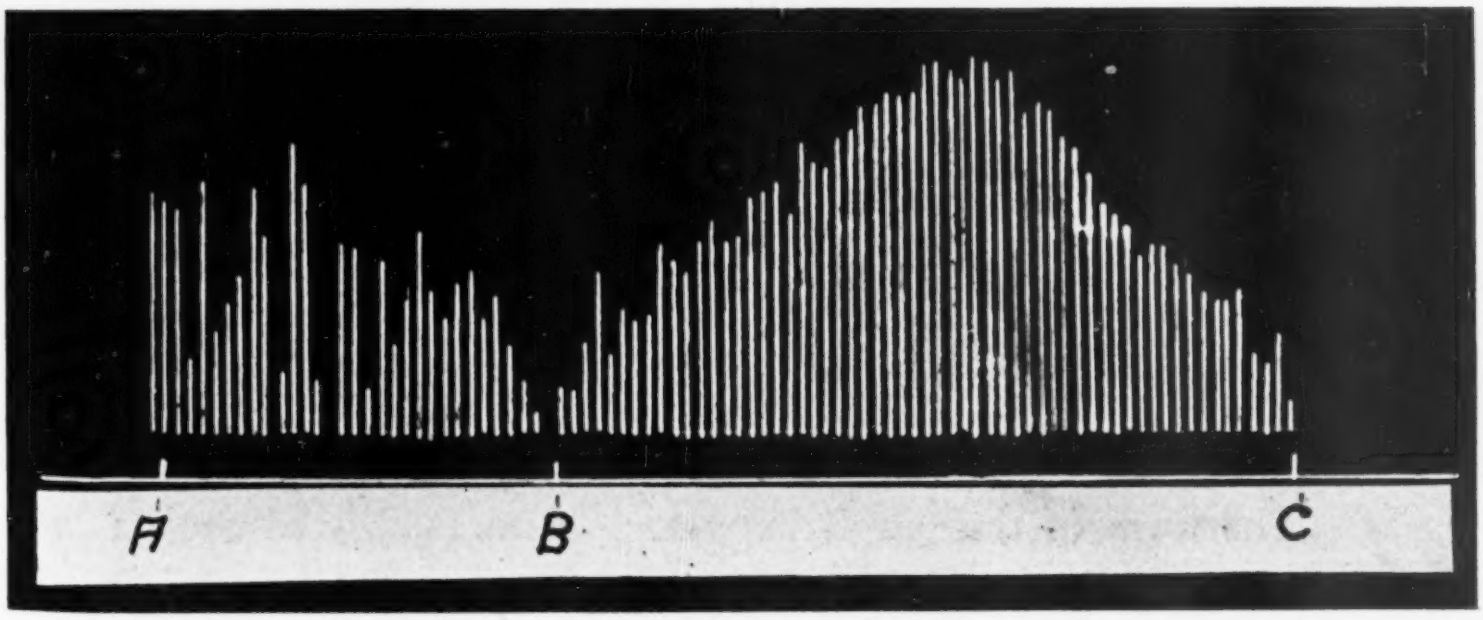

FIG. 7.[1] *A–B*, Conversation Period; *B–C*, Solution of an Original Problem in Machine Design

Instructions were given previous to the experiment that the subject sit passively in the chair and at a given signal select some problem and solve it to his own satisfaction.

[1] This record reads from left to right. It was made during the early part of our work before the machine was changed to its present form.

As soon as the subject was connected to the apparatus and the machine started the operator engaged him in conversation during *A–B*. At *B* the signal was given for him to start the problem. The operator then left the room. The problem was one suggested by an instructor previous to the experiment and consisted of designing, without pencil and paper, a method for automatically producing either make or break shocks, as desired, for use in the physiological laboratories. This experiment was carried out during the early part of our work on the knee-jerk. The record lacks an initial passive period and a final passive period was lost because the operator returned to the room just as the subject had finished his problem, as indicated at *C*. Table VII. gives the details of this record. It shows that the average ratio of the kicks during the active period is 1.70 times the average kicks during the conversation period. The record is given principally because of its regularity and its probable bearing on the *form* of a normal attention curve. The results shown in Table VII. are not included in the summary of the active/passive ratios in Table VIII.

TABLE VII

Interval	Time	Total Height of Kicks in Mm.	Number of Stimuli	Average Height per Kick	
				Active	Conversation
A–B.....	4′ 43″	670	33		20.30
B–C.....	8′ 26″	2044	59	34.64	

34.64/20.30 = 1.70, active/conversation, ratio

The form of the graph between *B* and *C* showed such a characteristic representation of what may prove to be the optimum tonicity conditions during free and spontaneous mental work, or, in terms of attention, showed such a gradual and regular rise and fall that it was decided to secure a more carefully controlled record from this subject.

Unknown to the subject the following experiment was performed. He was told that he was to perform an experiment similar to the one he had previously completed. He

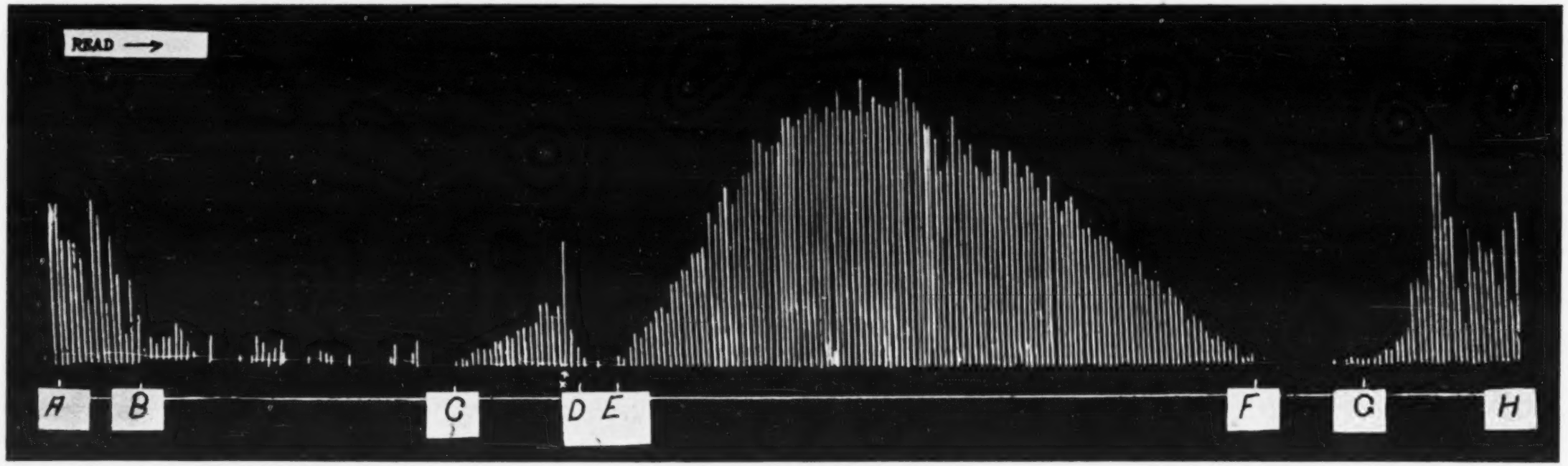

FIG. 7a.[2] *A–B*, Conversation Period; *B–C*, Passive Period; *C–D*, Active or Problem Period; *D–E*, Passive Period; *E–F*, Active or Problem Period; *F–G*, Passive Period; *G–H*, Conversation Period

[2] The graph reads from left to right.

was given as few instructions as seemed practicable to meet the conditions. During the course of the morning on which the experiment was performed a number of mechanical problems were brought to his attention. Among them was one which involved some difficulty in a wiring diagram. The data were presented to him orally but it so happened that at this time he was otherwise occupied and thus did not attempt a solution of the problem when he received the data. After he was connected to the apparatus, but before the machine was started, he was instructed to select a problem and solve it mentally, without pencil and paper, but not to start until the signal was given. During the experiment the operator sat in the room, behind the subject. No conversation was engaged in except as indicated. Fig. 7*a* is the record of his knee-jerk during two passive, two conversation and two active or problem periods.

When the machine was started ordinary conversation was kept up during *A–B* (Fig. 7*a*). At *B* the conversation ceased and the subject, with eyes closed, remained as passive as possible during *B–C*. At *C* the signal was given and he began his problem. At *X* the operator noted an unusually high kick and marked it on the record. He also observed that the kicks became extremely low, being absent at times during the interval *D–E*. At *E* the kicks returned, gradually increasing, reaching a maximum and then gradually falling to *F* where they disappeared. At *G* the operator engaged the subject in conversation, questioning him concerning the experiment. The subject reported that when the signal was given he began a problem which seemed to him too easy and hence he considered abandoning it. When questioned concerning the unusually high kick (*X*) he reported that he was aware of the kick and that it happened when he finally reached a definite decision to abandon the problem and start over. He further reported that at *E* he selected the problem concerning the wiring diagram previously mentioned. During *E–F* he solved it to his own satisfaction. In checking up on his solution afterwards he found it entirely adequate. During *F–G* the subject was as passive as possible, remaining thus

until interrupted by the operator. During this entire experiment the subject was free to work whatever problem he saw fit. The details of Fig. 7*a* are given in Table VII., *a*. It shows that: (1) the average height of the kicks during the active periods is 33.80 times the average height of the passive periods; (2) the average height of the kicks during the active periods is 1.22 times the average height of the kicks during the conversation periods; and (3) the average height of the kicks during the conversation periods is 27.61 times the average height of the passive periods. In the summary of Table VIII. only the active/passive ratio is included.

TABLE VII., *a*

Interval	Time	Total Height of Kicks in Mm.	Number of Stimuli	Average Height per Kick		
				Active	Passive	Conversation
A–B.....	2′ 9″	385	15			25.66
B–C.....	7′ 17″	114	51		2.23	
C–D.....	3′ 0″	280	21	13.33		
D–E.....	0′ 52″	4	6		0.66	
E–F.....	16′ 0″	6458	112	57.66		
F–G.....	3′ 26″	6	24		0.25	
G–H.....	3′ 52″	874	27			32.36
			Total.....	70.99	3.14	58.02
			Average.....	35.49	1.05	29.01

35.49/1.05 = 33.80, active/passive, ratio
35.49/29.01 = 1.22, active/passive, ratio
29.01/1.05 = 27.61, active/passive, ratio

The two active/conversation ratios of this subject seem to indicate some uniformity as shown by 1.70 in Table VII. and 1.22 in Table VII., *a*.

SUMMARY

Table VIII. is a summary of the data obtained from the subjects used in this experiment. The data presented in Table VII. are omitted as are also those presented in the column marked 'Conversation' in Table VII., *a*, since they do not fall under the definition which I have assigned to active and passive periods. Where more than one ratio for a subject is available these have been averaged before being included in the combined ratio for all subjects. From Table

VIII. we find from the ratio of the heights of the kicks of the active to the passive periods of the seven subjects considered that the knee-jerk tonus conditions under what we have designated as active or problem conditions is 9.91 times greater than under passive conditions.

Summary of Results

Table VIII

Subject Number	Table Number	Active/Passive Ratios	Combined Ratio for Subjects
1	1	10.13	10.13
2	2	2.40	2.56
	2*a*	2.73	
3	3	10.25	10.25
4	4	2.22	2.22
5	5	3.31	3.31
6	6	3.91	7.08
	6*a*	10.25	
7	7*a*	33.80	33.80
		Total.....	69.35
		Average.....	9.91.

Conclusions

The records from all subjects studied show that mental activity as defined in this paper increases muscle tonus as exhibited by the knee-jerk. In our experience no exception to this statement has been encountered. Furthermore, many incidental observations derived from a systematic series of knee-jerk experiments of different types confirm qualitatively the quantitative results of this investigation.

If mental activity of the problem type as used in this experiment involves the psychological factors of attention, then the conclusion seems justified that muscle tonus as measured by the knee-jerk is either (*a*) one of the factors in the attentive process, (*b*) a function of the attentive process or (*c*) the process of attention itself.

From graphs 7 and 7*a* the most probable form of the attention curve, when the problem is free and undirected, seems to be one in which there is a gradual rise in tonicity to a maximum, and then a gradual fall until the problem has been completed.

These data seem to demonstrate the truth of an assumption expressed by A. P. Weiss (5) that "what is psychologically known as attention is largely (*a*) an increased tonicity of the muscles which adjust the bodily posture for the favorable reception of the peripheral or internal stimuli to which the subject is responding; (*b*) an increased tonicity of the muscles through which the response is implicitly or overtly expressed."

Although no attempt has been made to select problems graded as to difficulty, the data herein presented seem to indicate that the greater the 'attentiveness' the greater the degree of tonus. This assumption seems justified by the fact that the solution of a problem evokes a greater degree of tonus than ordinary conversation, the relating of personal history, or receiving instructions.

All the records which present a period in which a single problem is solved tend to take the form of record 7 (*B–C*) and 7*a* (*E–F*). This, again, indicates that as attention approaches a maximum the tonus of the muscle increases proportionately; as the attention approaches the passive state the muscle tonus decreases proportionately.

It is planned to carry out further studies on this problem.

It is a pleasure to thank Professor R. G. Hoskins for directing this experiment and Professor A. P. Weiss who has read this manuscript and offered a number of valuable suggestions. I also wish to acknowledge the assistance of Mr. George Woodward who has assisted me throughout this work.

Bibliography

1. Lombard. The Variations of the Normal Knee-jerk, and their Relation to the Activity of the Central Nervous System. *Amer. Journ. Psychol.*, 1887, i, 2–71.
2. Howell. Text-book of Physiology, 8 ed., 1921, p. 158.
3. Tuttle, W. W. An Apparatus for Eliciting and Recording the Patellar Tendon Reflex. *Amer. Journ. Physiol.*, 1924, lxviii, 338–344.
4. Pressey and Tuttle. Senior Classification Test (Form *A*). Pub. School Pub. Co., 1922.
5. Weiss, A. P. Personal communication, 1924.

THE DEPENDENCE OF LEARNING AND RECALL UPON PRIOR MENTAL AND PHYSICAL CONDITIONS[1]

BY PAUL L. WHITELY

Introduction

This investigation is concerned with the dependence of learning and recall upon prior mental and physical conditions. The general problem resolves itself into three parts, each of which will be discussed separately in the interpretation of results. They are: (1) The effect upon learning of mental and physical activities introduced immediately preceding learning; (2) the effect upon subsequent recall of mental and physical activities introduced immediately preceding learning; and (3) the effect upon recall of mental and physical activities introduced immediately preceding recall. The physical activity in each case consisted of calisthenic exercises, lasting for a period of four or five minutes; whereas the mental activity consisted of simple multiplication exercises.

It is a well known fact that an individual's learning records are subject to wide daily fluctuations, and it is generally assumed that these variations are to a large extent due to differences of bodily and mental conditions and attitudes incident to the day's work. This experiment constitutes a part of a larger program designed to test the validity of this assumption by attempting to exercise some degree of control over the bodily and mental attitudes with which the subjects approach a given task.

The learning materials employed in this experiment were lists of three- and four-letter monosyllabic words. The words were either nouns or verbs, and many of the nouns could be construed as either nouns or verbs. Each list consisted of eighteen words, and there was the same number of four-letter and three-letter words in each list. Words were not repeated

[1] From the Psychological Laboratory of the University of Chicago.

again in succeeding lists except in two or three instances by mistake. There is no evidence, however, that this fact affected the results in any way. The words were selected so that they were all of about the same degree of familiarity, and none were used which are not in common usage. The words were carefully arranged, so that those which are obviously associated, such as hand and foot, did not occur together in the series. Words that rhymed or began with the same letter or ended with same letter did not occur together.

The words were serially exposed by means of a hand-operated memory drum. During learning a period of three seconds was given for each word exposure, while in recall a period of five seconds per word was permitted. The writer served as experimenter and practiced the operation of the drum until the process became more or less automatic. A watch placed on the table before the experimenter served as a check on the time, and it was found that after some practice the time interval could be guaged quite accurately.

The anticipation method was used in both learning and recall. The list of words was first presented once, and thereafter in all succeeding trials the subject attempted to anticipate each succeeding word just before it was presented. Each list was presented until the subject was able to anticipate successfully each succeeding word, when it was considered learned. The number of presentations for learning was taken to mean the total number of times each list was presented. A subject would learn list one, for example, and then recall it twenty-four hours later. On the third day, list two would be learned and recalled the fourth day, twenty-four hours later than the learning of list two. Most subjects came four times per week, though two or three came six times per week. From the nature of the experiment it will be observed that it was necessary to have them report an even number of times per week, *i.e.*, four or six times. The time of day for both learning and recall was kept constant.

Each subject was given preliminary practice lists in order that he might acquaint himself with the general method and technique of learning and also to obviate marked practice

effects. It was found that most subjects arrived at a practice level fairly rapidly. It did not appear necessary to give more than five preliminary practice lists to any of the subjects save one. This subject was given seven practice lists before beginning on the regular series. The results of the preliminary series were not used in computing the final averages.

Table I. represents a sample cycle which is inserted at this point to illustrate the order of procedure. An explana-

TABLE I

ILLUSTRATIVE CYCLE

Days	Subject				
	A	*B*	*C*	*D*	*E*
1.....	N-L-1	M-L-2	P-L-3	N-L-4	N-L-5
2.....	N-R-1	N-R-2	N-R-3	M-R-4	P-R-5
3.....	N-L-2	N-L-3	M-L-4	P-L-5	N-L-1
4.....	P-R-2	N-R-3	N-R-4	N-R-5	M-R-1
5.....	N-L-3	N-L-4	N-L-5	M-L-1	P-L-2
6.....	M-R-3	P-R-4	N-R-5	N-R-1	N-R-2
7.....	P-L-4	N-L-5	N-L-1	N-L-2	M-L-3
8.....	N-R-4	M-R-5	P-R-1	N-R-2	N-R-3
9.....	M-L-5	P-L-1	N-L-2	N-L-3	N-L-4
10.....	N-R-5	N-R-1	M-R-2	P-R-3	N-R-4

tion of the symbols used is as follows: L stands for learning and R for recall. The numbers following L or R indicate the number of the list of words. M refers to mental work, P to physical work, and N refers to the normal condition or no controlled activity preceding learning. Two or three specific illustrations will clarify the procedure more definitely. For example, subject *A*, on the first day, learned list 1 under the normal condition. The same subject recalled list 1 on the second day under the normal condition. Subject *B*, on the first day, learned list 2 preceded by mental work. The same subject recalled the same list on the second day under the normal condition, and so on. It may be seen from the illustrative cycle that each list was learned or recalled by each indivi-

dual under different conditions from that of every other individual. In other words, each list and each condition of learning and recall appear in all serial positions. This procedure eliminated the possibility of practice effects or differences of difficulty of the various lists favoring any one condition.

Three groups of five subjects each were employed. One group consisted of graduate students and these served throughout six experimental cycles. The remaining groups were composed of undergraduate students of psychology. One group completed two cycles and the other but one. For all conditions of learning and recall, directions were carefully prepared and typewritten on a card. Prior to performance each subject was given the directions to read, and quizzed briefly to make certain that they were correctly understood.

The physical work consisted of rather vigorous calisthenics for a period of four or five minutes. The window was opened and the subject was advised to take the exercises in the spirit of play. It was suggested that the exercises would be continued until the subject felt mentally and physically exhilarated. For the mental work, the subjects were required to perform simple multiplication tasks—multiplying two-place numbers. Thirty such problems were typewritten on a sheet —leaving sufficient room for computation. The subject was permitted to use any method he desired, and he was directed to solve the problems with the highest rate of speed consistent with accuracy. The purpose of the task was to key the subject up to a high degree of mental alertness.

Results and Discussion

1. *Effect upon Learning of Prior Mental and Physical Conditions*

The first problem is concerned with the effect upon the learning process of mental and physical activities introduced immediately preceding the event of learning. In this instance, learning under the normal condition is compared with the learning scores under the other two conditions. The method of comparison can best be seen by referring to the illustrative

cycle in Table I. Thus, to illustrate, conditions N-L-1, P-L-4, and M-L-5 were compared in the case of subject *A*. With subject *B*, the conditions N-L-3, P-L-1, and M-L-2 were compared, and so on.

Table II. shows the records for learning under the three conditions of experimentation. The column headed Normal indicates that no controlled condition preceded the learning activity. The column headed Physical signifies that physical work or calisthenics preceded learning; while the column headed Mental shows that mental work, or multiplication exercises, preceded learning. The numbers opposite the names of the subjects indicate the average number of trials per individual required to learn a list of words, and the numbers in parentheses signify the number of lists from which the average is computed. The table shows three sets of averages, namely, the average number of trials per individual, the average number of trials for each group, and the average of the total number of records.

Table II

Effect of Physical and Mental Work upon Learning

Subject	Normal	Physical	Mental
Ma	4.85 (6)	4.83 (6)	4.85 (6)
A	6.83 (6)	7.15 (6)	7.98 (6)
Mi	4.13 (6)	4.66 (6)	3.96 (6)
D	11.33 (6)	8.66 (6)	9.01 (6)
Hu	8.68 (6)	9.00 (6)	9.01 (6)
Average	7.16 (30)	6.86 (30)	6.96 (30)
C	6.00 (1)	10.00 (1)	8.00 (1)
Ha	13.00 (1)	18.00 (1)	14.00 (1)
G	7.00 (1)	8.00 (1)	9.00 (1)
Ho	5.00 (1)	5.00 (1)	3.00 (1)
K	8.00 (1)	6.00 (1)	9.00 (1)
Average	7.78 (5)	8.12 (5)	8.75 (5)
Ba	15.00 (2)	12.50 (2)	11.50 (2)
O'N	12.00 (2)	9.00 (2)	15.00 (2)
Od	5.00 (2)	8.00 (2)	8.00 (2)
Mar	11.50 (2)	15.50 (2)	12.50 (2)
Bl	6.00 (2)	6.00 (2)	6.00 (2)
Average	10.00 (10)	10.20 (10)	10.60 (10)
Total Average	7.64 (45)	7.89 (45)	7.95 (45)

An inspection of the tabular data reveals no significant differences for the three conditions of learning. For the graduate group the average for the normal condition is 7.16, for the physical condition 6.86, and 6.96 for the condition preceded by mental work. For the second group from whom only one record per individual was obtained, the averages for the three conditions of normal, physical work, and mental work are 7.78, 8.12, and 8.75 respectively; while the averages for the three conditions for the group from whom two records per individual were obtained are 10.00, 10.20, and 10.60 respectively. The differences between the averages for the three conditions are slight, and the records for the three groups are inconsistent. Evidently the two conditions are ineffective in influencing the group records.

TABLE III

CONSISTENCY OF INDIVIDUAL RECORDS OF LEARNING UNDER THE VARIOUS CONDITIONS

Subject	Physical		Mental	
	No. Cases Above Ave.	No. Cases Below Ave.	No. Cases Above Ave.	No. Cases Below Ave.
Ma.....	3	3	4	2
A.....	4	2	6	0
Mi.....	3	3	2	4
D.....	1	5	0	6
Hu.....	4	2	2	3

It is theoretically possible that the two factors may influence individuals differently, and hence the significance of the individual effects will be obscured in the group records. An analysis of the individual records of the graduate group for whom six measures for each experimental condition were obtained offers some support to this assumption. The values of Table III. indicate the number of times the records for the two controlled conditions were poorer or better than the average record for the normal condition. For subject *A* mental work was detrimental in all six cases, while it invariably exerted a beneficial effect in the case of subject *D*. Too few

records were obtained for any confident statements, but the data do suggest the possibility that the two conditions may affect some individuals and not others, while those affected may be influenced in either a beneficial or a deterimental manner.

2. *Effect of Mental and Physical Work upon Subsequent Recall*

In this phase of the problem, we were concerned with the possible effect which physical or mental work, introduced prior to the learning of any material, may exert upon the subsequent recall of that material. The subjects learned lists of words under the three conditions previously described and recalled them twenty-four hours later. Referring to the illustrative cycle of Table I., conditions N-R-1, N-R-4, and N-R-5, and conditions N-R-3, N-R-2, and N-R-1 were compared for subjects *A* and *B* respectively.

TABLE IV

EFFECT OF PHYSICAL AND MENTAL WORK BEFORE LEARNING UPON SUBSEQUENT RECALL

Subject	Normal	Physical	Mental
Ma	12.33 (6)	12.67 (6)	9.17 (6)
A	11.67 (6)	9.50 (6)	11.17 (5)
Mi	16.17 (6)	15.50 (6)	16.00 (6)
D	9.17 (6)	8.17 (6)	7.00 (5)
Hu	10.33 (6)	7.83 (6)	9.50 (6)
Average	11.93 (30)	10.75 (30)	10.68 (28)
C	17.00 (1)	18.00 (1)	15.00 (1)
K	12.00 (1)	6.00 (1)	14.00 (1)
Ha	5.00 (1)	12.00 (1)	4.00 (1)
G	9.00 (1)	—	—
Ho	12.00 (1)	13.00 (1)	13.00 (1)
Average	11.00 (5)	12.25 (4)	11.50 (4)
Ba	4.00 (2)	6.00 (2)	11.00 (2)
O'N	6.00 (2)	1.50 (2)	3.50 (2)
Od	11.50 (2)	14.00 (2)	13.00 (2)
Mar	11.00 (2)	12.00 (2)	13.00 (2)
Bl	13.50 (2)	6.00 (1)	13.50 (2)
Average	9.20 (10)	8.11 (9)	10.80 (10)
Total Average	11.22 (45)	10.40 (43)	10.80 (42)

TABLE V

CONSISTENCY OF INDIVIDUAL RECORDS OF SUBSEQUENT RECALL

Subject	Physical		Mental	
	No. Cases Above Ave.	No. Cases Below Ave.	No. Cases Above Ave.	No. Cases Below Ave.
Ma.....	3	3	1	5
A.....	2	4	2	3
Mi.....	2	4	4	2
D.....	2	4	0	5
Hu.....	0	6	1	5

The data are given in Tables IV. and V. The values of Table IV. indicate the average number of words recalled per list. Again no group effects are discernible. The differences are small and inconsistent from group to group. The values of Table V. indicates the possibility that certain individuals may be affected detrimentally. The records of subject Hu for physical work were below average in all six cases, while three subjects were affected detrimentally by mental work in five of six cases.

TABLE VI

EFFECT OF PHYSICAL AND MENTAL WORK UPON RECALL

Subject	Normal	Physical	Mental
Ma.....	12.33 (6)	12.66 (6)	13.50 (6)
A.....	11.67 (6)	11.33 (6)	13.80 (5)
Mi.....	16.17 (6)	15.16 (6)	15.65 (6)
D.....	9.17 (6)	9.36 (6)	10.40 (5)
Hu.....	10.35 (6)	9.50 (4)	7.16 (6)
Average.....	12.15 (30)	11.75 (28)	11.43 (28)
C.....	17.00 (1)	12.00 (1)	17.00 (1)
Ha.....	8.00 (1)	6.00 (1)	14.00 (1)
G.....	9.00 (1)	12.00 (1)	10.00 (1)
Ho.....	12.00 (1)	16.00 (1)	17.00 (1)
K.....	12.00 (1)	13.00 (1)	12.00 (1)
Average.....	11.37 (5)	12.25 (5)	13.25 (5)
Ba.....	4.00 (2)	7.00 (2)	11.00 (2)
O'N.....	6.00 (2)	2.50 (2)	2.00 (2)
Od.....	11.50 (2)	8.50 (2)	10.00 (2)
Mar.....	11.00 (2)	11.50 (2)	11.00 (2)
Bl.....	13.50 (2)	13.50 (2)	13.00 (2)
Average.....	9.20 (10)	10.60 (10)	9.30 (10)
Total Average.....	11.22 (45)	11.02 (43)	11.46 (43)

3. *Effect upon Recall of Prior Mental and Physical Work*

In this phase of the experiment the material was memorized under normal conditions, and the physical and mental work was introduced just prior to recall twenty-four hours later Referring to the illustrative cycle, conditions N-R-1, P-R-2, and M-R-3, and conditions N-R-3, P-R-4 and M-R-5 were compared for subjects *A* and *B* respectively.

The data are given in Tables VI. and VII. Again no group influence is manifested, while there is some indication that two of the subjects, *Mi* and *Hu*, may be detrimentally affected.

TABLE VII

CONSISTENCY OF INDIVIDUAL RECORDS OF RECALL UNDER THE THREE CONDITIONS

Subject	Physical		Mental	
	No. Cases Above Ave.	No. Cases Below Ave.	No. Cases Above Ave.	No. Cases Below Ave.
Ma.....	2	4	4	2
A.....	4	2	3	2
Mi.....	1	5	2	4
D.....	3	3	3	2
Hu.....	2	2	1	5

CONCLUSIONS

The group results are unambiguously negative in character. The data indicate quite conclusively that such amounts of physical or mental work preceding learning do not influence either the ability to learn a list of words nor the ability to recall such material twenty-four hours later. Neither does such work exert any deterimental effect when it is introduced prior to the recall of material that has been previously learned under ordinary conditions.

An analysis of the individual records of the graduate group suggests the possibility that a few individuals may be favoraby or unfavorably affected at times. Subjects *D* and *Hu* seem to be most susceptible to such influences. The records suggest that subject *D*'s ability to learn may have been favorably affected by both types of work, while mental work exerted a detrimental effect upon his ability to recall this material. Likewise *Hu*'s records seem to have been affected in three of the six conditions. However, no positive statements can be made on the basis of the number of individual records secured.

A STUDY OF ESTIMATES OF INTELLIGENCE FROM PHOTOGRAPHS

BY DONALD A. LAIRD AND HERMAN REMMERS

I. Introduction

Looks are not to be scoffed at. The interest in appearance, apart from styles, is very ancient. Although the soul may have been located in the viscera by the ancients the eyes were the windows of the soul, and mental qualities in general were thought to be reflected in the features of the face. That *some* insight may be gained into mental characteristics by the facial contour and expression is almost universal belief at the present day. This thesis is at the core of many of the widely advertised, and, we understand, widely sold systems that are issued in order to give the ordinary individual better insight into the inward character of his associates from their outward characteristics.

Regardless of whether the average-person-at-large is a disciple of Blackford, Freud, Dressler, Coue or any patent vendor they still possess a large modicum of belief in their own abilities in 'sizing up' others by appearances. The school superintendent demands a picture of the candidates for positions in the public schools. No picture, no job. A picture taken within certain date limits is one of the necessary qualifications for entering the consular service. And so the gauntlet runs from federal to individual in accepting their abilities in estimating certain characteristics from personal appearance, more especially from photographs.

II. Historical

Within fairly recent years it has come about that men have conscientiously attempted to check-up on their 'hunches' by experimental methods. Thus we find in the literature records of several noteworthy attempts to check-up on this common belief, that facial characteristics do betray the inner man.

These studies divide themselves into three groups: (1) Studies of the general nature of estimates, (2) Studies of the estimation of mental characteristics from physical characteristics, and (3) Studies of estimates from pictorial representations.

1. *Studies of the General Nature of Estimates.*—Hollingworth (1) has reported the most extensive studies in this field. These were concerned in the main with the agreement of a large group of judges in rating themselves and others in a list of somewhat abstract but still important traits. These traits were observed by first-hand acquaintances of the person rated, and not from photographs. Much of this work reported by Hollingworth duplicates that reported by Cogan (2). In general these two investigators have shown that in a group, of say twenty-four subjects, the average deviation of the estimates from the central tendency is about 4.4 places. In estimating intelligence, however, the A.D. is the smallest. Closest agreement was found in this trait, while beauty, apparently an objective affair, had an A.D. of 3.8. From their data it seems that in the case of desirable qualities the higher a person is rated in that trait the higher he stands in accuracy in estimating the possession of that trait by others. With undesirable mental characters the opposite holds. When the estimates of intelligence were checked up against rather makeshift mental tests and academic records a high positive correlation was found between these two criteria of intelligence.

The contributions of Norsworthy (3) and Cattell (4) also properly belong under this heading. The latter found that the agreement of scientific men regarding different traits of other scientific men varied, depending largely upon the trait estimated. Norsworthy, on the other hand, tested the consistency of various judges in their estimates from time to time. She did not find much disagreement between a first and a second estimation.

Self-estimates of the possession or degree of possession of certain mental characteristics makes another division of this work. The work of Hollingworth and Cogan, already cited, bears upon this topic. These investigators have found that

self-analysis is more variable than the estimates of associates, and tends on the whole to an overestimation of the desirable characteristics with a corresponding minimizing of the undesirable. The constant error in the self-estimation of intelligence was plus 3. Cattell found no tendency for scientific men to overestimate themselves, whatever tendency there was in that direction was balanced by counter tendencies to underestimation. Knight (5) worked out a study of a factor influencing self-estimates and found a tendency to place oneself nearer the ideal than the typical in estimating traits with a moral bearing.

The outstanding feature of all the studies reported so far is that the sole measure, with one exception, has been the disagreement in estimations, which really tells one little about the ability to estimate a given trait. It is much like the group of tenderfeet who look across the plains to their first mountain. The six persons estimate the distance as ranging from one mile to six, when in actuality it is at least twenty. Which is the more reliable measure: their deviation from the actual distance, or among themselves? It is the former measure that should be used. But in the case of the characteristics estimated in the studies so far reported there was only one instance in which there was some objective measure against which the estimations could be checked. This was intelligence, and admittedly the criteria used for that were not super-refined.

Some studies have been made, however, of the accuracy of estimates when checked up against standard measures. The earliest study was that reported by Simpson (6). He compared the estimates of intelligence of a group that had been given tests four years earlier, in 1907. His subjects were professors and students in Teachers' College. He obtained correlations that ranged from plus .97 to plus .65. On the whole the professors tended to have the higher correlations. Another early study was that made by Williams in California (7). He asked teachers in four cities in the state to select those pupils in their classes whom they regarded as feeble-minded. Twenty-six per cent. of the children selected by the

teachers as being of the feebleminded level of intelligence were in reality normal or superior when checked-up against the Binet scale.

Chassell and Chassell (8) made a somewhat similar study. They obtained a correlation of plus .42 between teachers' estimates of intelligence and the actual intelligence of the pupils in the first three grades as determined by the combined results on the group tests given. Fordyce, working in Nebraska, found a correlation of plus .44 between teachers' estimates of intelligence and actual intelligence as measured on standard scales. There was a tendency for the correlations to be higher in the higher grades. Working in the Saint Paul schools Varner (9) found that the teachers could estimate the upper and the lower twenty per cent. in intelligence with a fair approach to accuracy. The teachers were also able to estimate intelligence better than the I.Q. Lindsay (15), at the University Observational School, Iowa City, had a group of graduate students in education, their professor, and the teacher of the class estimate the intelligence of a group of tenth grade children. When correlated (*R*) with the Binet test the coefficients were found to range from plus .38 to plus .52. The lowest correlation was that of their regular teacher. All the visitors rated the intelligence of the pupils with greater accuracy than the teacher.

The experiment of Allport and Allport (10) stands in a separate class. They compared self-estimated intelligence with later measurements of intelligence. They found a tendency for those high in intelligence to underestimate their possession of the same while those low in the scores over-estimated; the correlation was minus .67.

The general results of this group of studies would rather discourage any conscientious attempt to estimate intelligence from "looks."

2. *Studies of the Estimation of Mental Characteristics from Physical Characteristics.*—Naccarati (11), presents a good historical survey of the work done in this field. He notes very diverse findings as the results of one investigator after another are passed in review. Most of this historical work is com-

paratively old; the measures of intelligence used consisted of school grades, or in some cases attendance at a truant school was taken as indicating inferior mentality.

A careful study of certain physical traits and the intelligence ratings of Columbia students revealed a plus correlation of .228 between the height-weight index and intelligence, while no significant correlation was found between height and intelligence; weight correlated negatively with intelligence. Using his morphologic index, obtained by dividing the length value of the limbs by volumetric value of the trunk, Naccarati found a slightly higher correlation than with the simple weight-height index.

Estimates of intelligence from behavior, unreliable as they have been found to be, are much more accurate than the physical characteristics as indicators of intellectual standing. When the two methods are combined, what should we expect? The historical work on this is included in the following section.

3. *Studies of Estimates from Pictorial Representations.*— The first work in this was done by Pintner (12). He used photographs of a group of twelve children, ranging from 5.7 to 12.5 years in mental age and with C.M.A.'s ranging from .47 to 1.93. Several groups of judges were used to arrange the photographs, physicians, psychologists, students, teachers, and a miscellaneous group. The pictures were arranged in the order of merit according to intelligence estimated by the judge. The correlations between estimated arrangement and the true arrangement ranged from minus .63 to plus .52. The highest correlation was made by a stenographer and the lowest by a psychologist! The median correlation was plus .10. With purely chance arrangements a correlation as high as a plus .43 was obtained.

It is quite apparent that these observers were not estimating the intelligence of these children as they arranged the pictures in order of merit. That what the subjects were estimating was much the same thing, however, is shown by the fact that the observers agreed among themselves to a very high degree in the arrangements. The correlations by groups for this ranges from plus .84 to a plus .94. In our opinion

Pintner passes this very significant fact over without sufficient mention. Although it was not intelligence as measured by standard tests that the judges were estimating, they still agreed closely in their estimates. Just what led the judges astray in this is a topic which we are studying for future publication.

Miss Pope (13) at Randolph-Macon Women's College used fifty photographs of seniors of that institution, all the likenesses being of women and all uniform to the extent that the subjects had worn cap and gown at the time the picture was taken. School ability was the criterion used for intelligence. The conclusions read: "the most outstanding of all the inferences is this that the average person has no basis whatsoever upon which to form any judgment of a stranger's intellectual capacity from a study of his features."

Anderson (14) undertook a somewhat similar study. He used sixty-nine photographs of department store employees as they appeared in the house journal. Twelve judges were used. The judges were told to pick out the seven most intelligent and the seven least intelligent faces from the group. Correlating only by groups in this way Anderson obtained a correlation of plus .27. Considerable individual variation in the correlations was an outstanding feature, even within the easy limits imposed in arranging by groups rather than *within* a group.

Hollingworth also reports (1) two experiments along this line. One, on the agreement of judges among themselves in rating pictures on intelligence, disclosed an A.D. of 2.86 when twenty photographs of women were used. In another, when the judgments of associates were checked up against the judgments of strangers from photographs the correlation on the trait of intelligence was plus .51 for all judges, using twenty-five photographs. When the judges were divided into groups, the *R* for women judges was found to be plus .61, and for a group of men judges plus .42.

III. Experimental

The Pictures.—Students enrolled in beginning psychology were requested to hand in to their instructor a photograph

taken within six months. The scores of these students in the Thorndike Intelligence Test for High School Graduates were available. From an assortment of about seventy pictures ten were selected in which both sexes were equally represented. All the pictures selected had as nearly as possible the same distance from chin to the hair-line on the forehead. All the pictures were retouched, all were as nearly full-face views as possible from the assortment. It was also attempted to have as even a gradation of Thorndike scores as possible.

Copies of the original photographs were made on uniform mountings, in uniform finish.[1] Readers who desire copies of these for class or personal use may secure sets through either of the experimenters.

The Judges.—These are described in detail in the report of the findings. In the main they consist of students of psychology in Colorado College, Purdue University, Yale University and of school superintendents. A considerable number of representative Main Street-Babbitt business men are also included among the judges.

The Method.—The pictures were arranged by all judges according to the order of merit method. A large group of the judges also arranged them by the method of paired comparisons as a check on the validity of the method. For a group of ten observers who made both arrangements in the same day an r of plus .901, with a P.E. of .013 was obtained between the rankings and the ranked scores on the Thorndike test. It was thought justifiable, therefore, to use the less laborious method. The well known formula

$$R = 1 - \frac{6\Sigma d^2}{N(N^2 - 1)}$$

was used in calculating the correlations between rankings by judges and rank in the test. The chief objection to this formula, *i.e.*, that it reduces to the Pearson r only when the distribution of ability is rectangular in shape, is met by the

[1] This work was done by Th. Winans, of the Department of Physics, of Colorado College. The pictures may be seen in the *American School Board Journal* for June, 1923.

fact that the scores of the ten individuals on the test do approximate such a distribution. Furthermore, since *N* in each correlation is 10, the additional statistical labor involved in obtaining the Pearson *r* is not warranted—particularly in view of the results obtained.

The Instructions.—The following instructions were on the outside of the envelope containing the pictures which was presented to the judges:

TEST DIRECTIONS: CRITERIA OF INTELLIGENCE

I

Spread the pictures out before you. Select the one that looks most 'intelligent' or 'brainy' to you and place it at your left. Next select the one that looks most stupid—*i.e.*, least 'brainy'—and place it at the extreme right. Then choose from the remaining eight the picture that seems to you to come second in 'intelligence' and place it next to the one at the left. In the same way choose the one next to the lowest in 'intelligence' and place it next to the one on the right. Continue in this manner until you have arranged all the pictures in a decending order of intelligence, the most intelligent looking face on the left, and the least intelligent looking face on the right.

Read no further until you have done this.

II

On a sheet of paper copy the numbers of the pictures in the exact order in which you have them ranked in intelligence, from left to right.

Now write as completely as possible what it was about each face that caused you to make the choices that you did.

On your record write your name, age, and occupation.

Thank you!

IV. FINDINGS

Estimates of Individual Judges

The pictures were arranged by several groups of individuals, each arranging the pictures without consultation with the others until after a satisfactory and final arrangement had been effected. Records were obtained in this way from sixty-four students in psychology at Colorado College, ten of whom were men; twelve students at the Sophie Newcombe Memorial College also arranged the pictures according to their estimate of intelligence. In addition to these arrangements were made by nineteen men, ranging from university president to a circus clown out of work during the winter, and by thirteen women ranging from doctors' of philosophy to charwomen.

Table I. shows the distribution of the correlation coef-

TABLE I

DISTRIBUTION OF CORRELATIONS OF ACTUAL ORDER WITH ESTIMATED ORDER BY JUDGES INDICATED

R	All Men	All Women	Students Only, Both Sexes	All Judges
−30 to −34	2	0	0	2
−29 −25	0	4	3	4
−24 −20	2	2	2	4
−19 −15	0	2	1	2
−14 −10	4	2	0	6
− 9 − 5	0	2	1	2
0 − 4	2	10	9	12
0 + 4	3	12	9	15
+ 5 + 9	0	0	0	0
+10 +14	3	3	2	6
+15 +19	5	11	9	16
+20 +24	2	4	5	6
+25 +29	2	3	2	5
+30 +34	0	3	2	3
+35 +39	0	0	0	0
+40 +44	1	4	2	5
+45 +49	0	3	3	3
+50 +54	2	2	2	4
+55 +59	0	3	3	3
+60 +64	0	4	4	4
+65 +69	0	0	0	0
+70 +74	0	3	3	3
+75 +79	0	1	1	1
+80 +84	0	0	0	0
+85 +89	1	1	1	2
Totals	29	79	64	108

ficients in the arrangement effected with the actual order of intelligence. The median correlation for the 108 individuals falls near + .15, with a range from − .33 to + .87. Inspection of the distribution table shows that there is no significant difference between any of the groups in their ability in estimating the intelligence of the subjects from their photographs. This ability, further, is for all practical and scientific purposes zero, although there are more positive coefficients than negative and although the negative coefficients are not of the same magnitude as are the positive ones.

Certain misplacements of some of the pictures which were very commonly made should be noted. Especially is this so in the case of pictures 1 and 10. In the case of the ratings

made by the Colorado College students photograph 1 was given its proper rank of first by only nine of the judges. It was given last rank by fifteen of the judges. Further, this same picture was ranked in the first half of the arrangement by only a third of the judges. These two photographs are not typical of all of the pictures used, but they serve to show very markedly how likely the chances are for great errors by the average judge in estimating relative intelligence within a group from appearances.

A detailed analysis of the ranks assigned to the different pictures by the Colorado students is given in Tables II. and III. These data may be used to determine the average position which may be assumed to have been given by these students by the method described in connection with the measurement of the relative connotative value of words (16). When this is done we find the following arrangement of the pictures expressing the tendency of this group of judges:

Picture Number	Actual Rank	Rank Assigned by Women Students	Rank Assigned by Men Students
1	1	8	10
2	8	7	6
3	7	5	2
4	6	6	9
5	9	4	7
6	4	3	5
7	2	9	4
8	3	2	3
9	5	1	1
10	10	10	8

These data will enable us to answer the question: Do women tend to overestimate the intelligence of women and men to overestimate the intelligence of men? Pictures 1, 3, 5, 6, and 9 are of women. Referring to the data immediately above we find that the women judges rated picture 3 two places higher than it belonged, picture 5 five places higher, picture six one place higher and picture 9 four places higher, a total misplacement by way of overestimation of 12 places. In the same way the overestimation of the pictures of the men by the women judges totals 11 places. This would indicate no favoritism due to the sex of the judge, when the

rankings of a large group of judges are pooled. The *R* between rankings by men and women is + 0.607.

TABLE II

TABULATION OF ESTIMATED ARRANGEMENTS BY WOMEN JUDGES SHOWING FREQUENCY OF RANKS ASSIGNED TO EACH PHOTOGRAPH. DATA FROM 54 COLLEGE WOMEN

Photograph Number	Rank Estimated in Intelligence										Totals
	1	2	3	4	5	6	7	8	9	10	
1	9	1	2	1	3	3	5	5	12	13	54
2	4	3	5	5	11	6	3	11	3	3	54
3	4	3	10	5	3	8	11	4	4	2	54
4	3	3	4	13	8	4	4	7	7	1	54
5	0	1	2	6	4	7	6	10	15	3	54
6	6	7	6	3	6	12	6	4	3	1	54
7	6	15	4	8	6	3	3	5	1	3	54
8	5	12	8	6	7	4	7	3	2	0	54
9	16	7	12	6	5	4	3	0	0	1	54
10	1	2	1	1	1	3	6	5	7	27	54
Totals.....	54	54	54	54	54	54	54	54	54	54	

On the other hand the pictures of men in this group were overestimated by the women by only 2 places and by the men to 4 places, indicating perhaps that in the pictures used favoritism is shown to the pictures of women by judges of both sexes.

TABLE III

TABULATION OF ESTIMATED ARRANGEMENTS BY MEN JUDGES SHOWING FREQUENCY OF RANKS ASSIGNED TO EACH PHOTOGRAPH. DATA FROM 10 COLLEGE MEN

Photograph Number	Rank Estimated in Intelligence										Totals
	1	2	3	4	5	6	7	8	9	10	
1	0	1	0	1	1	1	1	0	3	2	10
2	0	1	1	0	1	4	1	0	1	1	10
3	2	0	3	0	2	1	1	1	0	0	10
4	0	0	2	1	1	1	0	1	3	1	10
5	0	0	1	2	2	0	2	2	1	0	10
6	1	1	0	2	0	0	1	4	0	1	10
7	3	0	0	0	1	0	3	2	0	1	10
8	2	0	1	3	1	2	0	0	0	1	10
9	0	6	1	1	1	1	0	0	0	0	10
10	2	1	1	0	0	0	1	0	2	3	10
Totals.....	10	10	10	10	10	10	10	10	10	10	

The *index of brightness* (Otis Group Test, Form A) was available in the case of 61 of the judges. Correlating the I.B. with the success in approximating a correct arrangement of the pictures gave a coefficient of $-.14$. Possession of intelligence to a high degree does not predict greater success in the estimation of intelligence by this method than if a lesser degree or amount or kind, or whatever it is, of intelligence is possessed by the judge.

A study of the consistency of the judges who were very successful, or very unsuccessful, in their arrangement of the first set of pictures was made at this point. A second set of pictures was selected from the group originally handed in by the students. The pictures selected were all of women students. They did not approximate the uniformity of the first set. In this respect they may be considered very representative of the assortment that an employer receives from several candidates for a position.

Table V. shows the ability of the judges to rate this second set of pictures in comparison with their abilities indicated by the first set. Practically all the correlations obtained

TABLE V

CONSISTENCY OF ABILITY OF JUDGES IN RATING THE PICTURES

Sex of Judge	*R* on First Set of Pictures	*R* on Second Set of Pictures
Female	+.76	−.26
"	+.70	−.09
"	+.64	−.15
"	+.64	−.30
"	−.15	−.08
Male	+.52	+.04
"	−.21	−.15
"	−.33	−.50

with the second set are negative. This probably does not indicate a waning of the judges' ability so much as it does the fact that greater confusions in the judging are brought about by a marked lack of uniformity in the pictures being arranged. This is well shown in the case of pictures *I* and *J* in the second set. These are pictures of the *same subject.* One was taken for the high school annual and the pose reflects the solemnity

of the occasion. The other picture, taken a month later, is a typical 'sweetie' pose. The second picture is designated as *I*. Table VI. shows how these two pictures were placed

TABLE VI

PLACES IN ARRANGEMENT OF TEN PICTURES GIVEN TO TWO PICTURES OF SAME SUBJECT BY EIGHT JUDGES

Rank	Judge 1		2		3		4		5		6		7		8	
	I	*J*	*I*	*J*	*I*	*J*	*I*	*J*	*I*	*J*	*I*	*J*	*I*	*J*	*I*	*J*
1																
2	×									×		×				
3									×							
4															×	
5									×				×			×
6																
7						×										
8			×		×		×				×					
9		×														
10				×				×						×		
Places difference.	7		2		1		2		1		6		5		1	

by the judges in the consistency experiments. It will be noticed that in the case of over half of the judges the 'sweetie' pose was ranked the higher, although none of the judges suspected that they were two poses of the same subject. Pose *I* was overestimated 17 places and pose *J* only 8 places.

The findings relative to the ability of individuals to estimate rank in intelligence from the pictures may be summarized as follows:

1. With but few exceptions, which are probably due to chance, the ability to estimate relative position in intelligence from a series of photographs is largely zero.

2. There are no apparent sex or age differences in this. College students do no better or no worse than more mature persons.

3. There seems to be a tendency to overestimate the relative position of photographs of women, by both men and women judges.

4. There is no correlation between the possession of intelligence and the ability to rank the photographs.

5. Individual judges do not exhibit consistent ability in the ranking of pictures in this way from one set of pictures to another.

6. The ability to estimate relative intelligence from photographs is apparently lessened when the pictures are not fairly uniform.

Group Estimates

The next question to which an answer was sought was, Can *groups* of judges working as a committee obtain a higher correlation between their estimates and the true ranking than the single judge working alone? If so, does the size of the group have any bearing on the matter? Table VII. gives the data with reference to this part of the investigation.

Since school superintendents have often to hire teachers basing their judgment in part at least on a photograph of the applicant, it was thought desirable to test their ability with these ten photographs. An opportunity presented itself at the University of Iowa during the summer session of 1923. Notwithstanding the subjective assurance of some of the men as to their ability to judge intelligence in this way, it is evident from the table that superintendents are not significantly different from individual judges and grouped judges.

The remaining three categories of group judgments were obtained from students at Purdue University. The engineering students—62 in all—were all sophomores in the various engineering schools. The 'L.A.' students were students from the School of Science, which corresponds to the usual liberal arts college. Of these students there were 84, bringing the total of grouped judges up to 186.

It is evident from these data that the two questions we have essayed to answer must both be answered unfavorably. A group of judges working together can not arrange the ten pictures with any more accuracy than can a single judge. It also makes little difference whether the group consists of two, four, or five judges. With groups, as with individuals arranging the pictures, the relationship between rank in the intelligence test and rank assumed by the judges from the pictures is for all practical and scientific purposes neglible.

TABLE VII

DISTRIBUTIONS OF CORRELATIONS OF ACTUAL ORDER WITH ESTIMATED ORDER BY JUDGES INDICATED

R	School Supts., Groups of 5	Soph. Engineers, Groups of 2	Soph. Engineers, Groups of 4	L.A. Students, Groups of 4	All Groups
−.74 to −.70...	0	0	1	0	1
−.64 −.60...	0	2	0	0	2
−.44 −.40...	1	0	0	0	1
−.39 −.35...	0	0	1	2	3
−.34 −.30...	0	0	0	2	2
−.29 −.25...	0	2	0	0	2
−.24 −.20...	0	0	1	0	1
−.19 −.15...	1	3	0	0	4
−.14 −.10...	0	1	1	5	7
−.09 −.05...	0	1	0	1	2
−.04 .00...	0	1	2	1	4
0 +.04...	1	1	0	3	5
+.05 +.09...	0	2	1	0	3
+.10 +.14...	1	1	0	1	3
+.15 +.19...	1	0	0	2	3
+.20 +.24...	2	0	0	0	2
+.25 +.29...	1	0	0	0	1
+.30 +.34...	0	1	1	3	5
+.35 +.39...	0	0	0	1	1
Total groups....	8	15	8	21	52

Inspection of the original data for the groups (not reproduced here) shows the same misplacement of the pictures that has previously been noted. This is particularly true, again, of picture 1.

Estimates from a Series of Only Two Photographs

Eight pictures were taken from the set of ten that had been used and made into pairs. Two of these pairs had one picture of each sex, a third set was made of the pictures of two men, and a fourth set of the pictures of two women. There were three reasons for using only two pictures. First, there is a possibility that a large number is confusing in making it difficult for the judge to compare each picture with each other. Second, in practical situations where the candidates have been eliminated until only two remain the final selection is made from a pair. Third, it is possible to determine very easily when only two pictures are used whether or not the judges' selections are better, or as good as chance.

In Table VIII. data is given for the judgments of the intelligence from the pairs of pictures used. These show very clearly that the selections are no better than chance. Out of 292 judgments half of them are wrong. In none of the four sets of pictures used in this experiment was the accuracy of the judges any greater than is due to change. A selection with the eyes closed would have been as accurate.

TABLE VIII

RIGHT JUDGMENTS IN ARRANGING TWO PICTURES

Pair	Per cent. Right	Per cent. Wrong	Number of Judges
(*a*) Both men	46	54	84
(*b*) Both women	49	51	84
(*c*) One man, one woman	50	50	62
(*d*) One man, one woman	56	44	62
Average	50	50	

Sixty-two of the judges had arranged all four sets of the pairs of pictures. From the laws of chance one would expect about half of them to have made two selections properly, just as out of 62 pitches of a penny approximately half will be heads. Only one of the judges had all four selections right, 9 had three right, 38—slightly more than half—had half of the pairs of pictures right, 12 had only one pair in its proper arrangement, while 2 of the judges had every pair wrong.

We have found in the previous section that increasing the size of the groups of judges estimating, the relative intelligence of ten persons from their pictures does not increase the rightness of the estimates. Decreasing the number of pictures has been found not to increase the accuracy of individual judges.

V. SUMMARY AND CONCLUSIONS

Data have been gathered from 376 individuals with respect to their abilities in arranging persons by intelligence as judged from fairly uniform photographs. Objective measures of intelligence had been obtained from the Thorndike

Intelligence test for High School Graduates. The pictures were all of college students, such as a school board, or the civil service, or any prospective employer receives from his applicants.

It has been concluded that:

1. The individuals attempting to arrange the pictures of ten persons according to intelligence could have done as well with their eyes closed as with them open;

2. That any success better than this is due to luck, as was shown by having those who did well on one set of ten arrange another set, whereupon their apparent better abilities disappeared;

3. That women are no better at judging relative intelligence from photographs than are men;

4. That older persons have no better abilities in this than do the younger;

5. That the more intelligent persons have no better abilities in this than persons with less intelligence;

6. That there was a slight tendency for both men and women to overestimate the intelligence of the women from their photographs, this being, however, very slight;

7. That the ability to estimate intelligence in this way is lessened when the pictures are not uniform;

8. That in judging only two pictures, as in judging ten, one might as well close his eyes;

9. That a group of two, four, or five judges working together does no better than a single judge arranging the pictures by himself;

10. That one professional 'character reader and vocational expert' did no better than the average of the 376 persons studied in arranging these pictures.

References Cited

1. Hollingworth, H. L. Vocational Psychology. New York: Appleton, 1916. Pp. xviii + 308.
2. Cogan, L. G., et al. Self-analysis, Estimates of Associates, and Psychologica Tests. *School and Society*, 1915, *2*, 171–179.
3. Norsworthy, N. The Validity of Judgments of Character. In "Essays in Honor of William James."
4. Cattell, J. McK. Homo Scientificus Americanus. *Science*, *17*,

5. KNIGHT, F. B. Pitfalls in Rating Schemes. *Jour. Educ. Psychol.*, 1922, *13*, 204–213.
6. SIMPSON, B. R. Reliability of Estimates of General Intelligence with Applications to Appointments to Positions. *Jour. Educ. Psychol.*, 1915, *6*, 211–220.
7. WILLIAMS. Classification of Fifty Backward and Feeble-minded School Children. *Psychol. Clinic*, 1915, *9*, 97–106.
8. CHASSELL, C. F., AND CHASSELL, L. M. A Survey of the Three First Grades of the Horace Mann School by Means of Psychological Tests and Teachers' Estimates and a Statistical Evaluation of the Methods Employed. *Jour. Educ. Psychol.*, 1921, *12*, 243–252.
9. VARNER. Can Teachers Select Bright and Dull Pupils? *Jour. Educ. Research*, 1922, *6*, 126–133.
10. ALLPORT, F. H., AND ALLPORT, G. W. Personality Traits, Their Classification and measurement. *Jour. Abn. Psychol.*, 1921, *16*, 6–40.
11. NACCARATI, S. The Morphologic Aspect of Intelligence. *Arch. of Psychol.*, No. 45, 1921.
12. PINTNER, R. Intelligence and Estimations from Photographs. *Psychol. Rev.*, 1918, *25*, 286–296.
13. POPE, D. V. The Interpretation of the Human Face from Photographs. *Bull. Randolph-Macon Woman's College*, July–Sept., 1922, pp. 3–17.
14. ANDERSON, L. D. Estimating Intelligence by Means of Printed Photographs. *Jour. Applied Psychol.*, 1921, *5*, 152–156.
15. LINDSAY, E. E. Personal Judgments. *Jour. Educ. Psychol.*, 1921, *12*, 413.
16. LAIRD, D. A. Connotations as a Factor in the Mental Health of the Community. *Amer. Jour. Psychiat.*, 1921, *1*, 41–68.

THE CURVE OF LEARNING IN TYPESETTING

BY CHALICE M. KELLEY AND H. A. CARR

University of Chicago

With the already extensive literature on the subject of learning and the vast multitude of learning curves, one should hesitate before adding another one to the list. The study of the curve of learning in typesetting, or hand composition, is justified for the following reasons: (1) In the first place it is concerned with the study of an entirely different activity from any of those previously studied. (2) It is unlike most studies of learning in that it was made under actual school conditions. The majority of the studies have been made in the laboratory under experimental conditions and relatively few investigators have used classes doing regular school work. Thurstone,[1] in determining the learning curve equation, used classes in typewriting in a business school as did Chapman [2] in his study of positive acceleration. (3) This curve is unique in that it is based upon the entire work record of the subjects and not upon tests given at regular intervals or upon representative samples of the work done. By avoiding tests we eliminate any possibility of a test incentive and obtain a true record of the rate of learning. Tests naturally excite the individual to greater exertion than ordinary conditions of work and the result is a record of his maximum efficiency at the time. A curve plotted from tests given at regular intervals represents the greatest possible achievement at those time intervals and not the average rate of learning. (4) The practical value of such a curve is its fourth justification. Since it was obtained under actual trade school conditions it may be used as a norm of what may be expected of students in a similar situation.

[1] Thurstone, L. L., 'The Learning Curve Equation,' *Psychological Monographs*, Vol. XXVI., 1919.

[2] Chapman, J. Crosby, 'The Learning Curve in Typewriting,' *Journal of Applied Psychology*, Vol. 3, 1919, p. 252.

Typesetting, when learned, is as automatic and mechanical a process as is typewriting. The various characters or types, including capital and small letters, marks of punctuation, and the several sizes of spaces and quads, are placed in separate compartments in a standardized case. The copy which the compositor follows is placed on the edge of the case. The pieces of type are picked up separately with the right hand and placed in the composing stick which is held in the left hand. Each type has one or more nicks on one side which with the character on the upper surface makes it easy for the typesetter to place it right side up without the aid of vision. Each line is read over as it is finished and any errors that have occurred are corrected before the next line is set above it. Each line is properly spaced and justified as it is set. Spacing out a line to the full measure of the stick and justifying it so that none of the pieces of type will be loose enough to move out of position when it is locked up is one of the most difficult features to master. Correct spacing is extremely important and it varies from line to line according to the number of words per line, the height of the letters at the ends of the words, and the length of the line. At first, judgment enters into this part of the work but it later becomes almost as automatic as the rest.

Learning to set type, then, involves a mastery of the following factors. The first is the proper position with regard to the case and the correct manner of holding the composing stick. The second is becoming so familiar with the spatial arrangement of the characters in the case that the sight or thought of any given letter will cause the typesetter to reach automatically to that compartment in which those characters are placed. Next he must recognize by sight or touch the position of the character and the nick in order to place it in the proper position in the stick. The typesetter must learn to read directly from the type, which, since the letters are upside down, is the inverse of the printed page. The compositor must master correct spacing and justification. There is little uniformity in the finer details of setting type. Some compositors look at each piece of type as they pick it up

while others watch the copy closely and do not look at the type until they have set a full line, when they read it over to correct the errors.

The material for this study was collected at the School for Apprentices of the Lakeside Press, Chicago, from January to October, 1922. This is a trade school operated by the R. R. Donnelley Sons, Printers, to train employees for all departments of the concern. The training program requires that the boys who serve as apprentices spend half of each day in the school and the remainder of the time doing some work in the plant which for the beginners is chiefly errand service. The time in the school is divided between hand composition or typesetting and academic work related to the printing trade. Hand composition is part of the required preliminary course which every apprentice must take before he is allowed to specialize in any field. The course in typesetting consists of a standard series of seventy jobs or lessons, covering all of the essential phases of hand composition. The jobs are arranged in order of difficulty and all apprentices are required to complete these jobs in their regular sequence. The first jobs are of simple material known in trade parlance as straight matter, as contrasted with display matter, tabular matter, ads or job matter. The latter part of the course is given over to the more complicated phases of typography.

After a preliminary survey of the situation, it was decided to limit our study to the first twenty jobs for two reasons. (1) The later material was not homogeneous in character and it was doubtful whether any error or time values were comparable from job to job. (2) The students are often permitted to do extra jobs outside of the regular course after they have finished the first twenty jobs and consequently in this case it would be impossible to obtain any accurate measurement of the amount of practice. By limiting our study to the first twenty jobs, we are certain that all individuals in the group have had the same amount and kind of practice.

We obtained three sorts of data for each job: Its length, the number of errors made, and the time required to complete

it. The length of each job is measured in terms of the number of ems which is the typographical unit of measurement. An em is the square of the body of a type in any style or size. For example, in 12 point type it is a mass 12 points square (a point is 1/72 of an inch). An em is approximately the size of the body of the capital M in most styles and sizes of type. The number of ems thus represents approximately the number of separate movements required to set up a given job. The number of errors per job was obtained from the first proof sheets, which were preserved for this purpose. The time records for each student were furnished by the instructor, who kept a record of the time each job was begun and the time it was completed.

We were able to secure records for 51 individuals who completed the series of jobs in the regular order and who had no previous experience in typesetting before entering the school. We were not able to secure a complete record for all twenty jobs in every case. The students often failed to preserve the proofs and turn them in to the office as directed. The time records were occasionally missing, owing to the carelessness of the boy about having the time checked when the job was taken out or turned in, or to the absence of the instructor at the time the job was completed. Only individuals for whom we were able to secure at least five perfect records were included in our group. It thus follows that the data for the various jobs will be based upon the records of slightly different groups. This fact, however, does not militate against the validity of our curve, for the values for each point of the curve represent the average attainment of a representative group, all members of which have had the same amount and kind of previous practice.

This group of students was also quite homogeneous as to age and academic training. With two exceptions all individuals were sixteen or seventeen years of age. All were grammar school graduates. Only one fourth had had any high school training and but four of these had been in high school for more than a year. All members of the group, with five exceptions, were of foreign extraction, and the majority of them were Poles and Bohemians.

These records for the various jobs are given in Table I. The first column gives the number of the job and the second column states the length of each job in ems. The third column gives the number of individual records which were obtained for each job. The average time required to complete each job is given in column four, while the final column lists the average number of errors made per job.

TABLE I

Job. No.	No. of Ems Per Job	No. of Records Per Job	Ave. Time Per Job in Hours	Ave. No. of Errors Per Job
1	900	27	12.66	21.54
2	700	26	11.42	17.23
3	650	24	7.00	15.39
4	700	27	6.00	16.13
5	300	23	2.50	6.92
6	400	29	4.33	10.66
7	600	33	4.75	11.88
8	400	26	3.33	10.18
9	250	26	2.66	8.58
10	250	28	1.66	6.65
11	750	26	7.25	22.68
12	750	26	6.16	14.77
13	450	26	2.92	6.10
14	800	23	6.00	18.79
15	450	20	3.25	10.43
16	450	23	4.42	11.11
17	500	21	3.42	7.94
18	200	21	1.42	4.50
19	175	18	1.23	3.10
20	1300	17	8.43	16.56

The values representing speed of performance, amount of practice, and accuracy of performance may be derived from the data of Table I. The amount of practice was measured in two ways: In terms of the amount of time expended, and in terms of the amount of work performed stated in number of ems. These two series of values, representing the increasing amounts of practice from job to job, are given in columns 1 and 2 of Table II. The speed of performance for any job was also measured in two ways—in terms of the time required to set 1,000 ems of type, and in terms of the average number of ems set per hour. These two series of values representing variations of speed of performance from job to

job, are given in columns 3 and 4 of Table II. Any measure of accuracy must consider the total number of errors made in relation to the total number of errors which it is possible to make. Accordingly, we computed the total number of errors which it was possible to make, and then computed the average number of errors made for each 10,000 possibilities. These two sets of values for each job are listed in columns 5 and 6 of Table II.

TABLE II.

Job	Amount of Practice		Speed of Performance		Accuracy	
	Ems	Hours	Hours per 1000 Ems	Ems per Hour	Total No. of Errors Possible	Errors per per 10,000 Possibilities
1	900	12.66	14.01	72.48	9,714	22
2	1,600	24.08	16.27	61.35	7,519	23
3	2,250	31.08	10.77	92.59	7,337	21
4	2,950	37.08	8.61	116.27	6,474	25
5	3,250	39.58	8.70	114.93	2,514	28
6	3,650	43.91	10.93	91.74	5,035	21
7	4,250	48.66	7.92	126.55	5,655	21
8	4,650	51.99	8.33	120.46	3,505	29
9	4,900	54.65	10.52	95.24	3,564	24
10	5,150	56.31	6.60	151.51	2,129	31
11	5,900	63.56	9.61	104.16	9,435	21
12	6,650	69.72	8.21	121.95	9,582	15
13	7,100	72.64	6.42	156.25	3,518	25
14	7,900	78.64	7.49	133.33	7,710	24
15	8,350	81.89	7.20	138.88	5,285	23
16	8,800	86.31	9.92	101.01	6,642	17
17	9,300	89.73	6.86	147.06	6,936	11
18	9,500	91.15	7.10	140.84	2,812	16
19	9,675	92.40	6.97	144.91	1,980	16
20	10,975	100.73	6.41	156.25	10,130	16

From these values, six learning curves were plotted, four representing improvement in speed and two improvement in accuracy. Two typical curves are given in Figs. 1 and 2. Both curves have been smoothed to some extent by combining the data for the shorter jobs in such a way that the values represent the progress attained for approximately each successive 12.5 hours of practice.

There was a slow but uniform improvement in speed throughout the period of practice, while the most rapid rate of improvement occurred during the first thirty hours. The

initial high value is probably due to the fact that the subjects received considerably more help and supervision at this time than they did for the subsequent jobs. In general the speed curves were similar to the usual type of learning curve.

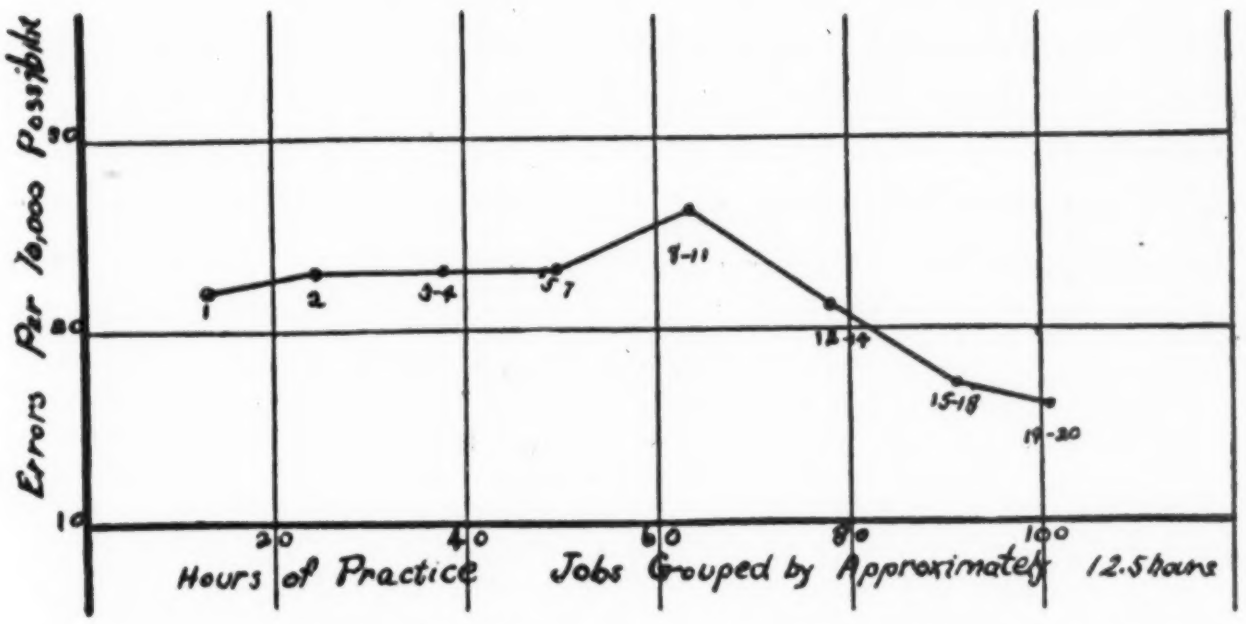

FIG. 1. Curve of accuracy. Number of errors made per 10,000 possibilities is plotted against amount of practice measured in terms of hours.

The error curves are unique, differing considerably from any that have hitherto been obtained. The amount of error gradually increased during the first twelve jobs, and this is followed with a final period of quite rapid improvement during

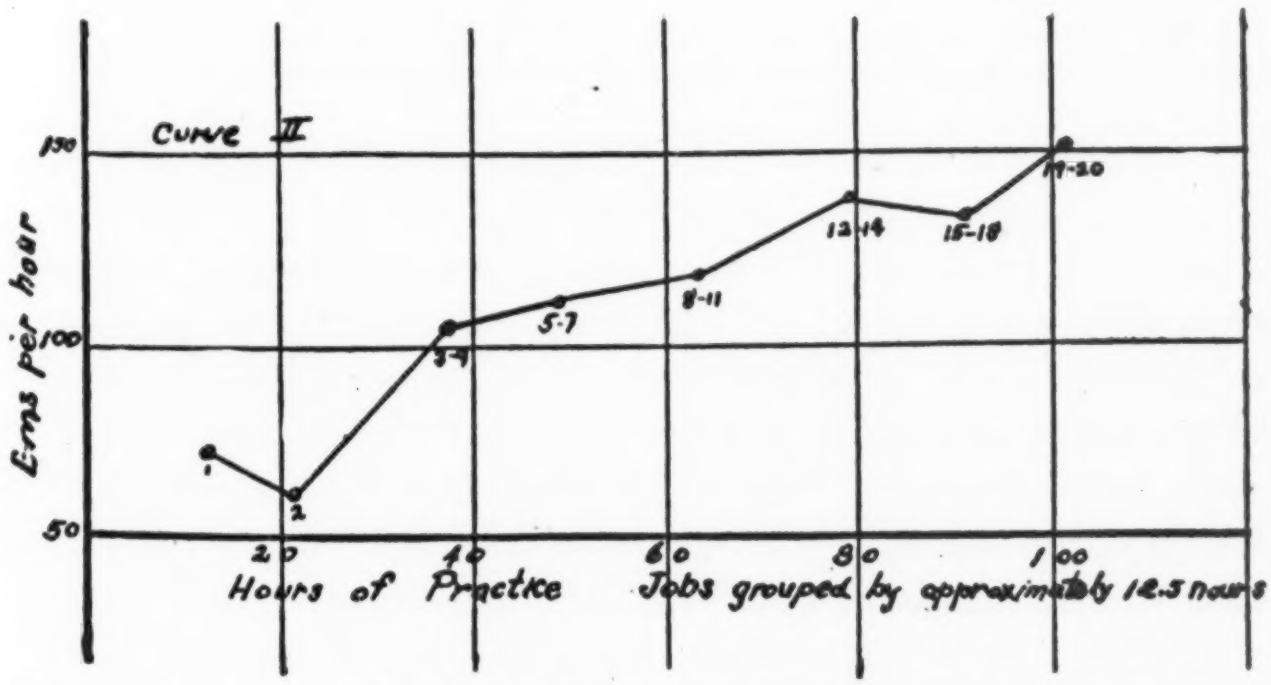

FIG. 2. Speed curve. Number of ems set per hour is plotted against amount of practice measured in terms of hours.

the latter part of the practice. The lack of initial improvement is probably due to the character of the material.

The errors were grouped in twelve classes. We computed the total number of possibilities for each kind of error and the average number of each made per individual. From these data we obtained the average number of each kind of error made for each 10,000 possibilities. The data are given in Table III.

The most frequent error in absolute terms is that of uneven spacing. Thirty-five per cent. of the total number of errors were of this kind. The four kinds of error made most frequently account for seventy-seven per cent. of the total.

TABLE III

Kind of Error	Ave. No. of Errors for 20 Jobs	Total No. of Errors Possible	No. of Errors per 10,000 Possibilities
Omissions	27.2	13,825	1.96
Wrong Paragraphing	5.93	80	7,410.00
Uneven Spacing	81.14	2,937	276.00
Wrong Division	10.60	314	337.00
Quads Showing Dim Letters	16.04	16,252	9.30
Wrong Capitals	.32	240	13.00
Spelling Substitutions	32.63	13,825	23.50
Wrong Font	38.39	13,825	27.70
Inverted Letters	6.00	13,825	4.30
Transpositions	7.38	13,825	5.30
Insertions	3.34	13,825	2.40
Wrong Alignment	3.28	13,825	2.30

In relative terms the most frequent error is that of wrong paragraphing—ninety-one per cent. of the total being of this kind. Three kinds of error—wrong paragraphing, wrong division and uneven spacing—account for ninety-nine per cent. of the total. The few possibilities presented for making these three kinds of error (see Table III.) may possibly account in part for the relative frequency with which they were made. Being encountered so infrequently, they will escape the attention of the students and but little opportunity will be presented for their elimination by practice. Because of their relative frequency these three kinds of error must have largely determined the form of the error curve as given in Fig. 1.

The error curves for paragraphing and uneven spacing were plotted separately. In both cases the values decreased slightly with but little variation with the exception of some *extremely high* records for certain jobs. In spacing these high values were obtained for jobs 3 and 4 and in paragraphing for jobs 10, 11, 13 and 15. Evidently these values largely account for the gradual rise of the error curve up to the eleventh job.

ON THE LOSS OF RELIABILITY IN RATINGS DUE TO COARSENESS OF THE SCALE

BY PERCIVAL M. SYMONDS

Teachers College, Columbia University

Freyd ('23) in his excellent article on the Graphic Rating Scale claims as an advantage for the scale that "the rater can make as fine a discrimination of merit as he chooses." In his own use of the scale Freyd employs a stencil divided into 20 units by which he transmutes the graphic ratings into scores. Although he states it nowhere definitely, there is the implication that by making the scale less "coarse" there is an increase in accuracy. Cady ('23), using the graphic rating device uses a scale of twenty-two intervals because "this method permits more elasticity of placement and readier translation into statistical values than that of two or three points as deviation from a mean."

Rating scales using preformed scale divisions vary all the way from two scale divisions up. Rating scales with only two steps have usually assumed the presence or absence of the trait. An odd number of intervals has generally been chosen, for no other reason that the writer can think of than that one usually wants an average position, disliking to make the judgment as to being better or worse than the average in close decisions. Galton ('83) used a nine point scale, Pearson ('07) used a scale of seven classes, Webb ('15) used a scale of seven classes, Downey ('19) used a scale of eleven classes, Plant ('22) made use of a ten scale class, Mendenhall ('22) has constructed rating scales with six to eight classes respectively, the Army Rating Scale contained five classes. Apparently the construction of rating scales has proceeded quite without consideration as to the reason for constructing scales with one rather than another number of classes.

This matter hinges on the statistical question as to how much reliability is lowered due to "coarseness of grouping."

Naturally one wishes in any measurement to use as fine a scale as possible, but just as in physical measurements it is useless to use a scale finer than the limits of one's eyesight or finer than the definiteness of the object to be measured, so in psychological ratings it is useless to use a scale finer than the judge's ability to discriminate or the definiteness of the trait in the subject. The degree of fineness to which this discrimination can be made in psychological ratings is determined by the coefficient of reliability. One has to choose how much of a loss in reliability one will permit in order to make easier the rating.

One will tolerate more of a loss in reliability when the reliability is low to start with than when the reliability is high. It is very much easier to raise a reliability of .30 to .31 than of .90 to .91. Every gain in reliability above .90 is very valuable. The real criterion for this is the coefficient of alienation [1] which indicates improvement in estimate over a random estimate. This improvement is rapid as correlation coefficients go above .90. Let us assume that for the purposes of rating we are willing to lose a reliability of not more than .01 when the true reliability is .91. This arbitrary unit is reasonable, for generally we are satisfied for ordinary purposes with reliabilities over .90. This corresponds to a change of coefficient of alienation of .0213. Hence a drop of reliability from .91 to .90 is equivalent to a drop in reliability (assuming the change in alienation to be identical)

from .205	to .00
.228	to .10
.285	to .20
.361	to .30
.446	to .40
.535	to .50
.627	to .60
.721	to .70
.816	to .80
.910	to .90
.9567	to .95

[1] Kelley, 'Statistical Methods,' pp. 173–174.

The formula for correcting a coefficient of correlation for coarseness of grouping is

$$r_{12} = \frac{p_{11}}{\sqrt{s_1^2 - \frac{h^2}{12}}\sqrt{s_2^2 - \frac{k^2}{12}}}$$

(Kelley: formula No. 100), where $p = \Sigma xy/N$ (which to a first approximation does not need correction for coarseness of grouping), h = the size of the interval of the first measure in terms of the units used, k = the size of the interval of the second measure in terms of the units used.

The difference between the true reliability and the lower one due to coarseness of grouping is

$$\frac{p_{11}}{\sigma_1\sigma_2} - \frac{p_{11}}{s_1 s_2} \equiv \frac{p_{11}}{\sqrt{s_1^2 - \frac{h^2}{12}}\sqrt{s_2^2 - \frac{k^2}{12}}} - \frac{p_{11}}{s_1 s_2}.$$

Let $h = k$ and let $s_1 = s_2 = 1$.

$$\text{Diff.} = p_{11}\left(\frac{1}{1 - \frac{h^2}{12}} - 1\right) = p_{11}\left(\frac{h^2}{12 - h^2}\right).$$

Then the problem becomes one of solving for h, the size of the class interval in terms of the standard deviation of all the cases being rated. For instance let us take the difference = .01 which was our arbitrarily assigned difference when the reliability is allowed to drop from .91 to .90. Since $s_1 = s_2 = 1$, $p_{11} = .90$, in this case

$$.01 = .90\left(\frac{h^2}{12 - h^2}\right),$$

$$h = .364.$$

When $p_{11} = .80$, diff. = .016

$$.016 = .80\left(\frac{h^2}{12 - h^2}\right),$$

$$h = .485,$$

and so on. Assume that the distribution to be rated covers a range of 5 σ. This is the most probable value for a population of 40.[1] Then dividing 5 by .364, the value of h obtained

[1] Kelley, *op. cit.*, p. 107.

above, gives 13.7, which is the number of class intervals necessary in order to have no more of a loss than .01 when the true reliability is .91. Carrying out this process one obtains the following table which shows the number of class intervals necessary in order not to effect a loss in reliability equivalent to a gain in the coefficient of alienation of .0213:

True Value of Reliability	Obtained Value of Reliability	Number of Class Intervals Necessary
.9567	.95	17.2
.910	.90	13.7
.816	.80	10.3
.721	.70	8.47
.627	.60	6.95
.535	.50	5.65
.446	.40	4.50
.361	.30	3.52
.285	.20	2.89
.228	.10	1.93
.205	.00	1.44

Webb ('15) finds the average reliability of a large number of ratings to be .55. By coincidence Voelker ('21) reports an average reliability of ratings at .55. Of course the reliabilities varied from very low to very high but with reasonable care, reasonable sagacity and rather close acquaintance, .55 is a good average figure for the ratings of traits of personality. From inspection of the above table it will be seen that a reliability of .60 may be obtained with a 7-class rating scale with no more than our arbitrarily defined loss of reliability. I have recently obtained character ratings from teachers who took only a moderate amount of interest in the task,—the average reliability was + .30. For such ratings a four-point scale is sufficient. Again I obtained ratings in a class in educational psychology on the difficulty of words in spelling and found an average reliability of + .90. In this case the task was of such a nature that a rating scale of 14 intervals would have been useful. But in general in the rating of human traits one cannot expect even under the best conditions reliabilities of over .60 or .70. In such rating there is no object in having rating scales with more than 7, 8 or 9 classes.

The graphic rating scale, although it may allow as fine a discrimination as the rater chooses to make, does not force such discrimination, and under ordinary rating conditions there is no advantage in obtaining more than a certain discrimination (as shown in the last table). The graphic rating scale may be desirable because of its convenience, simplicity, because it is pleasing to use, because it makes use of definitive verbal phrases, etc. But it contributes little to increased reliability.

This matter also has its application to mental and scholastic testing. A test with a reliability of .91 in a certain group may have scores ranging over a hundred points. This gives the illusory impression that one is getting some hundred degrees of ability. We have long known that a score of 48 on a mental test has a superiority over a score of 47 with only slight probability. But from the point of view of this article it is possible to say that a test with a reliability of .91 with a range of scores of a hundred or more points suffers the loss of only .01 in the coefficient of reliability were it to make discriminations of only 14 levels of ability. We have been using instruments graduated more finely than their accuracy in use warrants.

Conclusions

In constructing scales for rating traits of personality the optimum number of class intervals is 7. Rating with a scale of more than this number of classes demands a discrimination which does not yield an increase in reliability sufficiently great to make the increase worth while according to an arbitrary definition of such worth whileness. Likewise, according to our definition, a rating scale with a fewer number of classes suffers from a loss of reliability greater than is allowed, due to coarseness of grouping.

The graphic rating scale, although permitting as close a discrimination as possible, contributes but little to increased reliability over a 7-point scale.

Group mental and scholastic tests which cover a range of score of a hundred or more points give an illusory impression of discrimination.

Bibliography

Cady, V. M. ('23). The Estimation of Juvenile Incorrigibility. *Journal of Delinquency Monograph*, No. 2, Apr., 1923.

Character Education Methods ('22). The Iowa Plan $20,000 Award 1922. Character Education Institute, Chevy Chase, Washington, D. C.

Downey, J. ('19). The Will Profile: A Tentative Scale for Measurement of Volitional Pattern. *Univ. of Wyoming Bull.*, 1919, 16, 1–40.

Downey, J. ('23). The Will-Temperament and Its Testing.

Freyd, M. ('23). The Graphic Rating Scale. *Journal of Ed. Psychol.*, 14: 2, Feb., 1923, pp. 83–102.

Galton, F. ('83). Inquiries into Human Faculty.

Kelley, T. L. ('23). Statistical Methods.

Pearson, K. ('07). On the Relationship of Intelligence to Size and Shape of Head, and to Other Physical and Mental Characters. *Biometrika*, Vol. 5, pp. 105–146.

Plant, J. S. ('22). Rating Scheme for Conduct. *Amer. Journ. of Psychiatry*, Vol. 1, 1922, pp. 547–572.

Voelker, Paul F. ('21). The Function of Ideals and Attitudes in Social Education; An Experimental Study. Teachers College Contributions to Education, No. 112.

Webb, E. ('15). Character and Intelligence. *British Jour. of Psychol. Monographs*, No. 3, 1915.

LEARNING IN THE CASE OF THREE DISSIMILAR MENTAL FUNCTIONS

BY FOWLER D. BROOKS

The Johns Hopkins University

What are some of the differences between the learning curves of different mental functions? What relations exist between different kinds of learning, and how permanent is each kind?

Twenty-four graduate and undergraduate students in Johns Hopkins University practiced fifteen minutes per day for fourteen days on each of the following: cancellation of a's and t's, mental multiplication of three-place by two-place numbers, and inverted writing.

The score in cancellation is the number of a's and t's crossed out, minus the a's and t's omitted and other letters cancelled, the remainder being divided by fifteen.

In mental multiplication the multiplicands contained no 1's in ten's or hundred's place, and no 1's, 2's, 3's or 5's in unit's place. The multipliers contained no 1's or 2's in ten's place, and no 1's or 5's in unit's place. Such multipliers as 32 and 33 were omitted. All common short cuts were ruled out: *e.g.*, $489 \times 67 = \frac{67000}{2} - 67 \times 11$. The subjects had the problems before them as they worked. When the complete answer was obtained it was written down. No other use of pencil was allowed. The score is the number of correct figures in the answers, minus one tenth the number of incorrect figures.

For the inverted writing page one of the Courtis silent reading tests was used, the S's starting at the beginning of it each day. The S's were told to visualize the words as they would appear if they wrote them, to imagine them hinged at the bottom and turned down, and then to write them as they would appear. They were allowed to invert their papers

five times during each practice period and look through to see how the writing looked right side up. The score is the number of letters written per minute, minus corrections for such errors as omitting or incorrectly writing letters or parts of letters.

In Table I. are given the means and S.D.'s of each function for each practice period. In Fig. 1 are shown the three learning curves. Individual and percentile curves, as well as the 25-, 50-, and 75-percentiles, and the Q's are here omitted. *A* and *B* in Plate I. are specimens of the normal handwriting of two subjects. A_1A_2 and B_1B_2 are specimens of their inverted writing on the first and last days of practice. The numbers under the letters show the rates in letters per minute.

TABLE I

Means AND S.D.'S FOR THREE MENTAL FUNCTIONS

Practice Period	Cancellation of a's and t's		Mental Multiplication		Inverted Writing	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
1	37.9	9.7	19.5	14.2	9.1	4.3
2	41.4	10.4	28.4	18.6	14.4	7.1
3	48.5	11.2	30.7	20.1	19.7	9.7
4	53.2	9.5	37.9	25.3	23.3	9.4
5	55.9	11.2	42.4	28.5	26.3	10.3
6	57.9	11.0	46.4	31.8	27.8	11.3
7	61.7	9.4	50.7	38.3	30.9	11.1
8	62.5	8.8	54.0	36.6	33.6	10.5
9	63.3	10.1	56.0	33.8	35.2	9.6
10	65.9	8.7	59.5	39.3	38.6	9.6
11	68.6	10.0	64.9	39.5	40.9	9.1
12	71.3	10.8	67.1	43.4	42.6	11.0
13	72.0	10.7	71.1	41.0	43.7	10.4
14	71.9	10.9	74.2	41.6	46.2	9.8

II

The three curves have common features of learning curves, such as rapid initial rise, and periods of little or no improvement. The individual curves are not, of course, as smooth as the group curves of Fig. 1. They are very irregular, especially in mental multiplication. Many of them in inverted writing, however, are quite regular, doubtless because it was new and more interesting than the cancellation, and

not as difficult as the multiplication. Every subject improved in each function. Periods of no improvement are found in all individual curves, except one in the writing. They are more numerous in cancellation and multiplication.

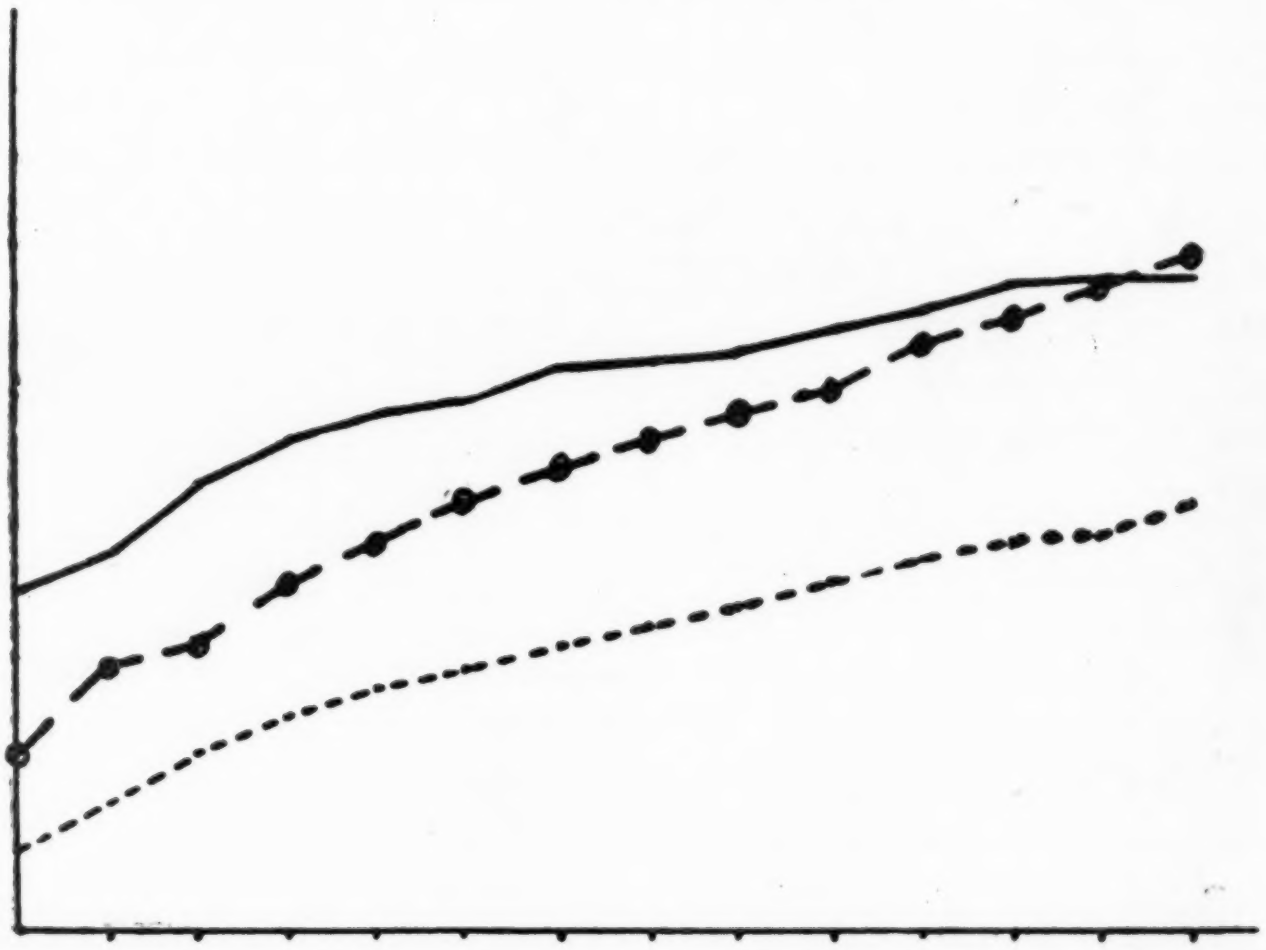

FIG. 1. Learning curves for three dissimilar functions.

Cancellation, a–t ————
Mental Multiplication ⊖ ⊖ ⊖
Inverted Writing – – – –

Positive acceleration after the seventh day of practice occurred in nine individual *c* [1] curves, in eleven *m* curves, and in fifteen *w* curves; they usually follow periods of little or no improvement.

Quality of inverted writing improved for all individuals. In most cases it was inferior to that of normal writing, but in a few cases it was as good or better. See *A* and A_2 in Plate I. Speed at the end of practice was from thirty to forty per cent. of speed of normal writing.

Improvement, expressed as mean gross scores, is greatest in *m* and least in *c*. The slope of the *m* curve is greatest;

[1] *c* means cancellation of a's and t's; *m* means mental multiplication; *w* means inverted writing.

The Kitten into Pleated Mean-Queen

A 1120 When the spring sun was warm
and the spring flowers had begun

To Little Life Forge Mary-Queen

B 100.8 spring flowers had begun to bloom
children far & away from their

The Kitten Who Played May-Queen

When the spring flowers had begun

And the spring sun was warm

Plate I

A
1170

When the spring flowers had begun

And the spring sun was warm

The Kitten Who Played May-Queen

B
8.001

When the spring sun was warm and the

spring flowers had begun to bloom

our children had a May-party

A1
14.7 When the spring sun was warm

The Kitten Who Played May-Queen

A2
470 spring flowers had begun to bloom our

When the spring sun was warm and the

The Kitten Who Played May-Queen

B1
1.2 The Kitten Who Played May-Queen

B2
32.9 When the spring sun was warm and the

The Kitten Who Played May-Queen

spring flowers had begun to bloom

C1
70 the

When the spring sun was warm and

C2
372 warm and the spring flowers

When the spring sun was

D1
138 warm and the spring flowers

When the spring sun was

D2
458 bloom, our children had a May-party

and the spring flowers had begun to

When the spring sun was warm

that of the *c* curve, least, although the *c* and *w* curves are very much alike. If improvement is expressed by the ratio of the final to the initial score, or as functions of the initial Q's or S.D.'s, it is greatest in *w*, and least in *c*. If it is expressed as functions of the average of initial and final Q.'s or S.D.'s, it is still greatest in *w*, but least in *m*.

The greatest improvement from this amount of practice is to be expected in inverted writing. The processes in a sensori-motor function such as cancellation,—attention to visual discrimination and simple motor reactions, are not as complex, and are used enough in everyday experiences to have the function in a relatively high state of efficiency. In mental multiplication the total process is much more complex, and greater possibilities of improvement are open than in *c*. Many of the basic connections are already formed, but are used in combinations different from those required in *m*. The new habits utilize many old processes and are built up rapidly. The great improvement in *w* is undoubtedly due to the utter newness of a relatively simple task.

Eight S's practiced one forty-five-minute period per day for fourteen days (one fifteen-minute period to each of *c*, *m*, and *w*, respectively, and in this order). Sixteen S's practiced fifteen minutes per day for forty-two days; the first fourteen days on *c*, the next on *m*, and the last on *w*. No significant differences appear. The curves for *c* and *w* are about the same for both groups, although those having successive practice made better progress in *w* after the fifth day. In *m* simultaneous practice gave a curve with steeper slope, especially during the last four days, but then this group made higher scores in *m* throughout the practice.

In *c* the quartile having the highest initial scores (Q_4) and the next two quartiles become more nearly alike, being nearly identical from the fourth to the twelfth day. Q_4 makes the least, and Q_1 the most improvement. In *m* Q_4 increases its lead over the other groups. The second and third groups keep about the same distance apart until the last three days when the second group (Q_{2-3}) makes much greater improvement than the group having the lowest initial scores. Q_4

makes the greatest, and Q_1 the least improvement. In inverted writing the initial differences between the highest quartile and the next two quartiles disappear. The quartile having the highest initial scores makes the least improvement, and the next two quartiles (Q_{2-3}) make the most.

The Pearson coefficients of correlation between initial score and amount of improvement are

$$r_c = -.474,$$
$$r_m = .517,$$
$$r_w = .029.$$

These data indicate that high initial ability in the more complex mental function is likely to be accompanied by greater improvement, whereas in the simple sensori-motor function of crossing out a's and t's it implies less improvement.

A positive relationship exists between improvement in cancellation and in inverted writing. The other relationships are probably neutral. The Pearson coefficients of correlation between improvement in the different pairs of functions are

$$r_{cm} = -.183,$$
$$r_{cw} = .414,$$
$$r_{mw} = .189.$$

The partials reveal what the totals suggest,—that improvement in cancellation and inverted writing has little in common with improvement in mental multiplication. The partial coefficients are

$$r_{cm \cdot w} = -.292,$$
$$r_{cw \cdot m} = .465,$$
$$r_{mw \cdot c} = .296.$$

The S's on the whole tend to maintain the same relative positions at the end as at the beginning of practice. The differences between beginning and final ranks are least in the highest or most complex mental function. The Pearson r's between initial and final scores are

	Raw Coefficients		Corrected Coefficients[1]
$r_{c_1c_{14}}$	.554 P.E.	.095	.616
$r_{m_1m_{14}}$	.744 "	.061	.812
$r_{w_1w_{14}}$	.470 "	.107	.652

At the beginning of practice the relative positions of the different subjects in cancellation and multiplication were somewhat the same, but the relationships between initial scores in *cw* and *mw* are probably neutral or slightly positive. At the end of practice the relative standings in the different functions are more nearly the same. The Pearson *r*'s are

Initial Scores		Final Scores	
$r_{c_1m_1}$	.481	$r_{c_{14}m_{14}}$	.507
$r_{c_1w_1}$	.184	$r_{c_{14}w_{14}}$	.402
$r_{m_1w_1}$	.190	$r_{m_{14}w_{14}}$	.262

Practice doubles the correlation between cancellation and inverted writing, raising it to a point which indicates a positive relationship between the two functions.

Practice increases the variability. The initial and final standard deviations and ranges are given in Table II. Here too it appears that practice differentiates individuals to the

TABLE II

INITIAL AND FINAL S.D.'S AND RANGES

	Initial S.D.	Final S.D.	Initial Range	Final Range
Cancellation	9.7	10.9	45.1	41.1
Mental Multiplication	14.2	41.6	68.1	171.0
Inverted Writing	4.3	9.8	16.0	37.4

greatest extent in the most complex function. Training for the same amount of time makes individual differences greater in all three functions, but it leaves them most alike in the relatively simple task of crossing out a's and t's, and makes them least alike in the complex task of mental multiplication.

After ten weeks of no practice the group was re-tested.

[1] Spearman's formula,

$$r_{xy} = \frac{\sqrt{r_{x_1y_2}\, r_{x_2y_1}}}{\sqrt{r_{x_1x_2}\, r_{y_1y_2}}}, \text{ was used.}$$

It lost 6.37 per cent. of its gain in cancellation, 23.38 per cent. of it in inverted writing, and 43.12 per cent. of it in mental multiplication. Individual changes ranged from a gain of 13 per cent. to a loss of 52 per cent. in *c*; from a gain of 11 per cent. to a loss of 102 per cent. in *m*; from a loss of 1 per cent. to 55 per cent. in *w*. The S making the gain in mental multiplication after ten weeks of no practice was worried by sickness at home during most of the practice days. Learning which involves the higher associative processes seems to be less permanent than that which involves a sensori-motor function.[1]

Conclusions.—A comprehensive comparison of plateaus in the three functions cannot be made because the learning could not be continued long enough. Practice seems to preserve the ranks of the individuals in the three functions, but at the same time it makes them less alike. Identical training, instead of reducing individual differences, increases them in such amounts in the three functions that the more significant differentiation of individuals seems to be in the higher intellectual traits. Improvement is more permanent in cancellation, and least permanent in mental multiplication. A high initial score implies greater improvement in the latter, but less in the former. Improvement in cancellation implies improvement in inverted writing.

[1] E. L. Thorndike, 'Educational Psychology," II., 304. J. B. Watson, 'Psychology from Standpoint of a Behaviorist,' p. 342f.

EFFECTS OF INTERVAL ON RECALL

BY WARNER BROWN

Psychological Laboratory, University of California

A list of common words was presented to be learned, and immediate recall was demanded. After intervals of eight or sixteen minutes a second recall was demanded without warning. After intervals of from three to seven days there was a third (unexpected) recall. Eighteen groups of subjects, in all 531, were employed.

The first 48 words of the Berrol key list were used. They were presented by the experimenter as follows: each word was read aloud, and then a brief phrase containing the word; finally the whole list was read through as a whole. As soon as the reading had been finished a written recall was demanded, for which five minutes' time was allowed. The students supposed this to be the end of the test, and then gave their attention to memorizing difficult lists of nonsense syllables. At the end of either eight minutes or sixteen minutes of work with the nonsense syllables, a second five minutes recall of the list of words was demanded. Data were obtained during the week from sections or groups of sections of an elementary course, and on Monday an unexpected five minute test (the third) was given to the whole class for retention of the words. Data are shown in Table I.

Effect of Interval between First and Second Recall.—Recitation after sixteen minutes is not much inferior to recitation after eight minutes. The longer interval is possibly better as measured by the effect upon retention after four or five days. The facts are shown in Table II. It will be noted that none of the differences are large enough to be considered reliable even though the number of cases is large.

Effect of Interval between Second and Third Recall.—As the interval between the time of learning and subsequent recall is increased from three to seven days, there is a slight falling

TABLE I

AVERAGE NUMBER OF WORDS RECALLED OUT OF 48, AND MEAN SQUARE DEVIATIONS

Group	Number of Cases	First Recall	S.D.	Second Recall 8 Min. after 1st.	S.D.	Second Recall 16 Min. after 1st.	S.D.	Third Recall	S.D.	Interval before Third Recall
1	26	21.8	4.4			19.4	5.3	17.6	5.6	3 days
2	30	21.2	6.0	20.4	7.1			17.9	6.7	3
3	41	22.9	4.8	22.0	5.6			19.2	5.4	3
4	25	22.6	4.5			21.6	5.3	19.4	4.6	3
5	28	21.9	4.0			20.2	5.7	16.6	5.5	3
6	19	22.3	5.4	21.0	5.7			16.9	7.0	4
7	25	21.3	4.6	21.1	5.3			16.4	5.0	4
8	27	21.3	4.5			18.9	5.0	16.3	4.6	4
9	29	20.6	5.4	19.8	5.0			16.7	5.4	4
10	42	24.1	5.6			22.5	6.3	18.5	5.6	5
11	30	22.9	4.3			22.1	4.3	16.5	3.9	5
12	13	20.5	5.8			20.3	6.4	14.6	6.0	6
13	29	21.9	4.1	21.2	4.7			15.8	4.8	6
14	54	22.5	5.2	20.9	5.2			16.3	5.8	6
15	47	20.6	5.0			19.6	5.6	15.0	5.5	6
16	11	19.3	5.2	18.1	4.2			13.2	5.2	7
17	29	19.8	4.5			17.4	5.3	17.1	6.9	7
18	26	21.8	5.1	20.5	5.8			13.3	4.6	7

off in average retention as measured by subtracting the amount recalled on test 3 from the amount retained at the time of learning (test 2).

TABLE II

DATA GROUPED ACCORDING TO INTERVAL BETWEEN FIRST AND SECOND RECALLS

	264 Cases, 8 Min. Interval	S.D.	267 Cases, 16 Min. Interval	S.D.	Difference in Favor of 8 Min. Interval	P.E. of Difference *
First recall ..	21.81	.319	21.84	.305	−.03	.297
Second recall.	20.79	.342	20.29	.350	.50	.330
Third recall .	16.54	.362	16.89	.343	−.35	.336

Table III. shows that there is a steady loss from the third to the sixth day, but there is reversal in the case of the seven day interval, and the loss from one day to the next is slight. The correlation between interval and amount forgotten computed, not from Table III., but from Table I., is .56 for first recall minus third and .47 for second minus third.

* Computed: P.E. $= .6745\sqrt{\sigma^2 M_1 + \sigma^2 M_2}$.

TABLE III

DATA GROUPED ACCORDING TO INTERVAL BETWEEN SECOND AND THIRD RECALLS, SHOWING DIFFERENCE BETWEEN AVERAGE OF RECALL 2 AND AVERAGE OF RECALL 3

Intervals Days	Number of Cases	Differences between Averages	P.E. of Difference *
3	150	2.62	.276
4	100	3.54	.325
5	72	4.62	.411
6	143	4.88	.341
7	66	3.76	.547

The Relation between Amount Learned and Amount Retained.—If we designate the three recalls by the numbers 1, 2 and 3, as in Table IV., then it does not appear that the correlations between 1 and 2 are higher when the interval between 1 and 2 is eight minutes than when it is sixteen minutes. It is, however, clear that the correlation between 1 and 3 is likely to be higher when the interval between 1 and 2 is eight minutes than when it is sixteen minutes.

TABLE IV

CORRELATIONS WITHIN EACH GROUP OF SUBJECTS BETWEEN RECALL 1 AND RECALL 2, AND BETWEEN RECALL 1 AND RECALL 3, SEGREGATED ACCORDING TO INTERVAL BETWEEN RECALL 1 AND RECALL 2, AND REARRANGED ACCORDING TO SIZE OF CORRELATION

Interval 8 Min.		Interval 16 Min.		Interval 8 Min.		Interval 16 Min.	
r_{12}	P.E.	r_{12}	P.E.	r_{13}	P.E.	r_{13}	P.E.
.96	.02	.90	.03	.87	.03	.75	.05
.93	.02	.90	.02	.82	.05	.73	.06
.91	.03	.86	.02	.76	.06	.72	.08
.89	.02	.86	.02	.75	.06	.72	.07
.86	.03	.83	.04	.75	.07	.65	.07
.86	.03	.82	.04	.72	.06	.58	.09
.84	.04	.79	.05	.70	.07	.57	.09
.81	.05	.78	.05	.69	.10	.54	.09
.68	.09	.74	.06	.64	.08	.52	.07
M = .86		M = .83		M = .74		M = .64	
Difference between means = .03 P.E., .022				Difference between means = .10 P.E., .024			

* Computed as follows: P.E. = $.6745\sqrt{\sigma^2 M_1 + \sigma^2 M_2 - 2r_{12}\sigma_1\sigma_2}$.

In my opinion a very important light is thrown by these data upon the entire discussion of the relation between learning and recall. In the data (see Table IV.) the correlation between the amount first learned, which is of course the inverse measure of ease of learning, and the amount retained (second or third recall), is found to be very close in some groups and far less close in other groups. The groups, however, are, many of them, as large or larger than the whole number of subjects employed in much quoted investigations of this subject; and the correlations obtained are, for many of the groups, highly reliable. When diverse results are obtained from groups of subjects under conditions which are as nearly as possible uniform it is not to be expected that different investigators working under different conditions will always be able to verify the existence of the relation. The evidence here indicates, however, that a positive relation is normal, and that when conditions are favorable the correlation is very close between amount learned and amount retained.

The relation between amount learned and amount retained gradually falls off with increase in the length of the interval between the learning and the final test of retention. This is true no matter whether the learning is measured by the first recall or the second, as can be seen in Table V. Although this decrease is not very large it is regular. The correlation between length of internal and size of r is $-.34$ for r_{13} and $-.53$ for r_{23}. These figures are computed, not from Table V., but from the r's for each of the 18 groups.

TABLE V

CORRELATIONS BETWEEN RECALL 1 AND RECALL 3, AND BETWEEN RECALL 2 AND RECALL 3, GROUPED ACCORDING TO THE INTERVAL BETWEEN RECALL 2 AND RECALL 3 AND SEVERAL CORRELATIONS AVERAGED FOR EACH INTERVAL

Interval, Days	r_{13}	r_{23}
3	.74	.83
4	.70	.82
5	.63	.78
6	.68	.72
7	.65	.71

Is the Amount Retained after Several Days More Closely Related to Original Learning (First Recall) or to Recitation (Second Recall)?—It is found that it is more closely related to recitation as may be seen in Table VI.

TABLE VI

CORRELATIONS AND PARTIALS BETWEEN RECALL 3 AND RECALL 1 OR 2; CALCULATED FOR EACH GROUP OF SUBJECTS AND AVERAGED FOR 9 GROUPS

	Av. 9 Correlations r_{31}	P.E. of Average	Av. 9 Correlations r_{32}	P.E. of Average	Av. 9 Partials $r_{31.2}$	P.E. of Average	Av. 9 Partials $r_{32.1}$	P.E. of Average
Interval between 1 and 2 was 8 min.....	.74	.02	.81	.02	.10	.01	.49	.03
Interval between 1 and 2 was 16 min.....	.64	.02	.75	.01	.05	.01	.51	.02

Examination by the method of partial correlations shows the relation more clearly. The correlations between recall 3 and recall 1 are seen to be low when recall 2 is held constant, and the correlations between recall 3 and recall 2 are relatively high when recall 1 is held constant. The conclusion is clear that what we will remember after three to seven days depends more upon what we can remember for eight or sixteen minutes than it does upon what we remember directly after learning. Retention depends upon successful recitation.

Reminiscence.—If the reader compares the data of this experiment with those of my experiment on 'reminiscence' (this JOURNAL, VI., p. 379, 1923) he will note that although the list used was the same and the methods of teaching nearly the same, the conditions were not such as to get a second or third recall better than the first. Averages are shown in Table VII.

TABLE VII

	First Recall	Second Recall	2 Minus 1
Former experiment, 30 minutes interval, 193 subjects...	25.5	26.8	1.3
Present experiment, 8 minutes interval, 264 subjects....	21.8	20.8	− 1.0
Present experiment, 16 minutes interval, 267 subjects...	21.8	20.3	− 1.5

No analysis has been made of the items as was done in the former experiment. I attribute the poorer absolute recall to poorer teaching (one repetition of each word was omitted as compared with the former method), and to poorer coöperation (the former subjects were members of my own class, the latter not). The fact that the second recalls are now slightly inferior to the first recalls, whereas they were in the other experiment slightly better, I attribute to the factors just mentioned, which affect all recalls unfavorably and further, probably, to retroactive inhibition from the very hard nonsense syllables memorized in the interval.

It may be noted that the method of correlation applied to this material confirms the writer's findings in the former study in so far as the third recall is shown to be closely connected with the success of the recitation in the second recall.

NOTE ON BUILDING LIKES AND DISLIKES IN CHILDREN

BY FRED A. MOSS

George Washington University

and

Bureau of Public Personnel Administration

With the exception of the experiments reported by Watson relatively little attention has been paid by psychologists to the conditioning of behavior in young children. And yet in these early experiences we have the genetic explanation of many peculiarities of later behavior which are generally looked upon as inherent idiosyncrasies. It is in a large measure to this early conditioning that we must look for a satisfactory explanation of such problems as why a person is very fond of one fruit, but can not stand another; why certain scenes and certain sounds are very pleasing to one person but cannot be endured by another; why a person may be very fond of one color, but think another hideous. That these likes and dislikes as well as many of our other later preferences and aversions are the direct result of early conditioned reflexes, seems to be supported by the following experiments.

Experiments with Children.—In the hope of gaining a more accurate understanding of how conditioned reactions are built up, I conducted a series of experiments with two children, Billie, age two, and Joe, age four. The object in taking children as young as these was to get back as close as possible to their original reactions and to responses affected very little, if any, by the so-called "purposeful" element.

Billie, Age Two.—The experimental conditions were as follows: Three cups were prepared, one filled with orange juice which had been made very sweet with sugar, another filled with vinegar which was very, very sour, and the third filled with pure water. A medicine dropper was provided for each liquid. The child was blindfolded and told to open his

mouth. A portion of orange juice was squirted into his mouth. He was told to open it again and again and each time for the first six times he was rewarded with orange juice. By this time the strangeness of being blindfolded and receiving orange juice in this unusual way had in a measure worn off and he was no longer excited. So when he opened his mouth the seventh time, I administered a large portion of vinegar. The effect was instantaneous. He began spitting, shook his head, gritted his teeth and seemed to shiver all over. He also began to cry, but by a speedy administration of orange juice his tears were soon dried and the investigation again proceeded. Orange juice was again given continuously for five times, but on the sixth time the vinegar was substituted. Simultaneously with the giving of the vinegar, I clicked a little telegraph snapper with which he had previously been playing and clicking himself. Practically the same reaction appeared this time as the first time the vinegar was given. The orange juice was again given four times followed by the vinegar and snapper. After repeating this ratio of orange juice and vinegar twice, Billie's patience was exhausted and the experiment had to be discontinued until the next day.

On the second day of the experiment the first conditioned reaction appeared. This was accomplished by clicking the snapper at the same time I administered a medicine dropper filled, not with vinegar, but with pure water. The conditioned reaction appeared twice on this day. On the third day it appeared twice, on the fourth day five times, and after that it gradually became more and more firmly established until in the end when the snapper sounded the same reaction almost invariably occurred regardless of whether the dose be vinegar, orange juice or water.

The blindfold was removed and after he had gone back to playing I clicked the snapper and got no response. It seemed, therefore, that the reaction was more or less conditioned by the whole situation of being blindfolded and receiving liquid into his mouth under such unusual conditions, and could not at that stage be set off by the sound of the snapper alone. However, after the experiment had been continued for twelve

days, I clicked the snapper while he was quietly playing by himself and the shivering reaction occurred. After that I tried him at various times and was often able to get the response by clicking the snapper when he was not noticing me. This would seem to indicate that conditioned reactions in their initial stages occur in response to a whole group of stimuli, and that it is only in their later stages that a single element is abstracted and can serve as an adequate stimulus.

Billie manifested an excellent example of conditioned likes and dislikes. Before the experiment he liked the snapper and often played with it. But after the experiment had been continued for two weeks he didn't like it and threw it down when it was placed in his hand. Before the experiment he was very fond of oranges, but subsequently he lost all his relish for oranges and very much preferred apples.

Joe, Age Four.—With Joe I had more difficulty in getting a conditioned reaction by means of the vinegar and snapper combination. In fact the vinegar did not seem to worry him very badly, and he would willingly sit the live-long day and take the vinegar or orange juice, provided I furnish enough candy, chewing gum or nickels to make it worth his time. Althouth I never succeeded in getting from him the shivering reaction, I did get the withdrawing and spitting occasionally when I substituted water or orange juice for the vinegar and clicked the snapper. The lack of success in his case appears to be very largely due to the fact that the vinegar did not affect him strongly. As the experiment went on he seemed to become negatively adjusted to the vinegar and would about as soon swallow it or hold it in his mouth as to spit it out.

An electric plate was later substituted for the liquids in the case of Joe. The usual procedure for developing a motor conditioned reflex was followed. He was given an electric shock simultaneously with the clicking of the snapper for seven times. The shock naturally caused the jerking of the hand away from the plate. The shock was then occasionally omitted until in a short time it was entirely eliminated, but the jerking of the hand occurred almost invariably at the

sound of the snapper, no matter what the child was doing when the snapper sounded.

In his later behavior he had not the strong dislike for oranges which Billie manifested, but he also preferred apples or bananas. He did show, however, the same dislike for the snapper as a plaything.